THE INFORMATION IN CONTINGENCY TABLES -
AN APPLICATION OF INFORMATION-THEORETIC CONCEPTS
TO THE ANALYSIS OF CONTINGENCY TABLES

by

C. T. Ireland and S. Kullback

Technical Report No. 235
August 4, 1976

Prepared under Contract N00014-76-C-0475
(NR-042-267)
Office of Naval Research

Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

The Information in Contingency Tables -
An Application of Information-Theoretic Concepts
to the Analysis of Contingency Tables

by

C. T. Ireland[1] and S. Kullback[2]

1. Introduction

The primary purpose of this paper is to present an exposition of the methodology underlying the analysis of the information in contingency tables. We shall stress the concepts, techniques, analyses and inferences without entering into extensive technical statistical proofs or detailed references to the bibliography at the end.

It is useful to note that we are concerned with an aspect of multivariate (multiple variates) analysis with particular application to qualitative or categorical as well as quantitative variables. The basic data we deal with are counts in multiway cross-classifications or multiway contingency tables. Multiway contingency tables, or cross-classifications of vectors of discrete random variables provide a useful approach to the analysis of multivariate discrete data.

As we shall see, the analytic procedures serve to bring out various interrelationships among the classificatory variables in a multiway cross-classification or contingency table in many dimensions. Classical problems in the historical development of the analysis of contingency

[1] C. T. Ireland is Professor of Statistics, George Washington University.
[2] S. Kullback is Visiting Professor of Statistics.

tables concerned themselves with such questions as the independence or conditional independence of the classificatory variables, or homogeneity or conditional homogeneity of the classificatory variables over time or space, for example. Such classical problems turn out to be special cases of the techniques we shall discuss. These techniques result in analyses which are essentially regression type analyses. As such they enable us to determine the relationship of one or more "dependent" qualitative or categorical variables of interest on a set of "independent" classificatory variables as well as the relative effects of changes in the "independent" variables on the "dependent" variables. In particular such problems as the determination of possible factors and measures of their effect and interactions in the representation of the logits of one or more dichotomous variables lend themselves to the analysis we shall examine.

The methodology is based on the Principle of Minimum Discrimination Information Estimation, associated statistics and Analyses of Information. General computer programs are available to provide the data for the inferences.

2. Contingency Tables

We assume that the reader has some familiarity with cross-classifications in the form of contingency tables. We use a slightly modified conventional notation. For example, for a four-way contingency table, that is, one with four classifications or variables, each of several categories, not necessarily the same in number, we represent the observed number of occurrences in the (ijkl) cell of the contingency table by x(ijkl), where the indices i,j,k,l, range over the respective categories of the variables. The corresponding probabilities are represented by p(ijkl). Summation over one or more indices, resulting in various marginal distributions or marginals, is indicated by a dot or dots, thus

$$\sum_{i} x(ijkl) = x(\cdot jkl), \quad \sum_{j}\sum_{l} x(ijkl) = x(i\cdot k\cdot), \text{ etc.,}$$

with a similar notation for the probabilities.

We shall denote estimates under various hypotheses or models by $x^*_\alpha(ijkl)$, where values of the subscript α will range over the hypotheses or models.

An example of a 2x2 two-way contingency table is shown in Table 2.1.

Table 2.1

	x(ij) j = 1	j = 2	
i = 1	x(11)	x(12)	x(1•)
i = 2	x(21)	x(22)	x(2•)
	x(•1)	x(•2)	x(••) = n

The estimated two-way table under the hypothesis or model of independence is shown in Table 2.2.

Table 2.2

	x*(ij) j = 1	j = 2	
i = 1	x(1•)x(•1)/n	x(1•)x(•2)/n	x(1•)
i = 2	x(2•)x(•1)/n	x(2•)x(•2)/n	x(2•)
	x(•1)	x(•2)	n

A common statistical measure of the association or interaction between the variables of a two-way 2x2 contingency table is the cross-product ratio, or its logarithm. The cross-product ratio is defined by

$$\frac{x(11)x(22)}{x(12)x(21)}\,, \tag{2.1}$$

though we shall be more concerned with its logarithm

$$\ell n\ \frac{x(11)x(22)}{x(12)x(21)}\,. \tag{2.2}$$

We shall use natural logarithms, that is, logarithms to the base e, rather than common logarithms to the base 10, because of the nature of the underlying mathematical statistical theory. Note that with the estimate for independence, or no association, the logarithm of the cross-product ratio is zero.

$$\ell n\ \frac{x^*(11)x^*(22)}{x^*(12)x^*(21)} = \ell n\ \frac{\frac{x(1\cdot)x(\cdot 1)}{n}\ \frac{x(2\cdot)x(\cdot 2)}{n}}{\frac{x(1\cdot)x(\cdot 2)}{n}\ \frac{x(2\cdot)x(\cdot 1)}{n}} = \ell n\ 1 = 0\,. \tag{2.3}$$

The logarithm of the cross-product ratio is positive if the odds satisfy the inequalities

$$\frac{x(11)}{x(21)} > \frac{x(12)}{x(22)} \quad \text{or} \quad \frac{x(11)}{x(12)} > \frac{x(21)}{x(22)}\,,$$

since then we get for the log-odds

$$\ell n\ \frac{x(11)x(22)}{x(12)x(21)} = \ell n\ \frac{x(11)}{x(21)} - \ell n\ \frac{x(12)}{x(22)} > 0$$

$$= \ell n\ \frac{x(11)}{x(12)} - \ell n\ \frac{x(21)}{x(22)} > 0\,.$$

The logarithm of the cross-product ratio is negative if the odds satisfy the inequalities

$$\frac{x(11)}{x(21)} < \frac{x(12)}{x(22)} \quad \text{or} \quad \frac{x(11)}{x(12)} < \frac{x(21)}{x(22)}\,,$$

since then we get for the log-odds

$$\ell n \frac{x(11)x(22)}{x(12)x(21)} = \ell n \frac{x(11)}{x(21)} - \ell n \frac{x(12)}{x(22)} = 0$$

$$= \ell n \frac{x(11)}{x(12)} - \ell n \frac{x(21)}{x(22)} = 0 .$$

The logarithm of the cross-product ratio thus varies from $-\infty$ to $+\infty$. Later we shall consider procedures for assessing the significance of the deviation of the logarithm of the cross-product ratio from zero, the value corresponding to no association or no interaction.

For the three-way 2x2x2 contingency table in addition to the classic types of independence, interaction or association, there arises an additional one important historically and practically. This is known as no three-factor or no second-order interaction. No three-factor or no second-order interaction implies that the logarithm of the association measured by the cross-product ratio for any two of the variables is the same for all the values of the third variable, that is, there is no second-order interaction if

$$\text{(2.4)} \qquad \begin{cases} \ell n \dfrac{x(111)x(221)}{x(121)x(211)} = \ell n \dfrac{x(112)x(222)}{x(122)x(212)} , & i, j \\ \ell n \dfrac{x(111)x(212)}{x(112)x(211)} = \ell n \dfrac{x(121)x(222)}{x(122)x(221)} , & i, k \\ \ell n \dfrac{x(111)x(122)}{x(112)x(121)} = \ell n \dfrac{x(211)x(222)}{x(212)x(221)} , & j, k . \end{cases}$$

One is concerned with the possible hypothesis or model of no second-order interaction when none of the other types of independence are found. However, in this case, the corresponding estimate cannot be expressed explicitly in terms of observed marginals although the estimate is constrained to have the same two-way marginals as the observed table. Straightforward iterative procedures exist to determine the estimate under the hypothesis or model of no second-order interaction. For the general three-way rxsxt contingency table there are of course many more relations among the log cross-product ratios like (2.4) which must be satisfied, but the iterative procedures to determine the estimate extend to the general case with no difficulty.

For four-way and higher order contingency tables the problem of presentation of the data increases, as do the variety and number of questions about relationships of possible interest and varieties of interaction. The basic ideas, concepts, notation and terminology we have mentioned for the two- and three-way contingency tables extend to the more general cases as we consider the methodology. For some additional prefatory remarks see Ku et al (1971).

3. Discrimination Information

To make the discussion more specific and with no essential restriction on the generality, we shall present it in terms of the analysis of four-way contingency tables. Let us consider the collection of four-way contingency tables RxSxTxU of dimension rxsxtxu . For convenience let us denote the aggregate of all cell identifications by Ω with individual cells identified by ω so that the generic variable is $\omega = (i,j,k,\ell)$, $i = 1,\ldots,r$, $j = 1,\ldots,s$, $k = 1,\ldots,t$, $\ell = 1,\ldots,u$. Suppose there are two probability distributions or contingency tables (we shall use these terms interchangeably) defined over the space Ω , say $p(\omega)$, $\pi(\omega)$, $\sum_{\Omega} p(\omega) = 1$, $\sum_{\Omega} \pi(\omega) = 1$. The discrimination information is defined by

$$(3.1) \qquad I(p:\pi) = \sum_{\Omega} p(\omega)\ \ln \frac{p(\omega)}{\pi(\omega)} \ .$$

The basis for this definition, its properties, and relation to other definitions of information measures will not be considered in detail in this exposition. For the particular types of application to which we shall restrict this exposition the π-distribution, $\pi(\omega)$, in the definition (3.1) according to the problem of interest may either be specified, or it may be an estimated distribution. The p-distribution, $p(\omega)$, in the definition (3.1) ranges over or is a member of a family of distributions of interest.

Of the various properties of $I(p:\pi)$ we mention in particular the fact that $I(p:\pi) > 0$ and $= 0$ if and only if $p(\omega) = \pi(\omega)$.

4. Minimum Discrimination Information Estimation

Many problems in the analysis of contingency tables may be characterized as estimating a distribution or contingency table subject to certain restraints and then comparing the estimated table with an observed table to determine whether the observed table satisfies a null hypothesis or model implied by the restraints. In accordance with the principle of minimum discrimination information estimation we determine that member of the collection or family of p-distributions satisfying the restraints which minimizes the discrimination information $I(p:\pi)$ over all members of the family of pertinent p-distributions. We denote the minimum discrimination information estimate by $p^*(\omega)$ so that

$$I(p^*:\pi) = \Sigma\ p^*(\omega)\ \ln \frac{p^*(\omega)}{\pi(\omega)} = \min I(p:\pi) \ . \tag{4.1}$$

Unless otherwise stated, the summation is over Ω which will be omitted.

In a wide class of problems which can be characterized as "smoothing" or fitting an observed contingency table the restraints specify that the estimated distribution or contingency table have some set of marginals which are the same as those of an observed contingency table. In such cases $\pi(\omega)$ is taken to be either the uniform distribution $\pi(ijk\ell) = 1/rstu$ or a distribution already estimated subject to restraints contained in and implied by the restraints under examination. The latter case includes the classical hypotheses of independence, conditional independence, homogeneity, conditional homogeneity and interaction, all of which can be considered as instances of generalized independence and will be considered in some detail in this paper. By generalized independence is meant the fact that the estimates may be expressed as a product of factors which are functions of appropriate marginals. See Ku et al (1971).

5. Minimum Discrimination Information Statistic

To test whether an observed contingency table is consistent with the null hypothesis or model as represented by the minimum discrimination information estimate we compute a measure of the deviation between the observed distribution and the appropriate estimate by the minimum discrimination information statistic. For notational convenience and later

computational convenience let us denote the estimated contingency table in terms of occurrences by $x^*(\omega) = np^*(\omega)$. For the "model fitting" or fitting class of problems, that is, with the restraints implied by a set of observed marginals (those of a generalized independence hypothesis), the minimum discrimination information (m.d.i.) statistic is

$$2I(x:x^*) = 2\Sigma\, x(\omega)\, \ln \frac{x(\omega)}{x^*(\omega)} \tag{5.1}$$

which is asymptotically distributed as a χ^2 with appropriate degrees of freedom under the null hypothesis.

The statistic in (5.1) is also minus twice the logarithm of the classic likelihood ratio statistic but this is not necessarily true for other kinds of applications of the general theory.

6. Minimum Discrimination Information Theorem

We now present a theorem which is the basis for the principle of minimum discrimination information estimation and its applications. We shall present it in a form related to the content of this discussion on the analysis of contingency tables.

Let us consider the space Ω mentioned in Section 3 and the discrimination information introduced in (3.1). Suppose now, for example, that there are three linearly independent statistics of interest defined over the space Ω,

$$T_1(\omega)\ ,\ T_2(\omega)\ ,\ T_3(\omega)\ . \tag{6.1}$$

Let us determine the value of $p(\omega)$ which minimizes the discrimination information

$$I(p:\pi) = \Sigma\, p(\omega)\, \ln \frac{p(\omega)}{\pi(\omega)} \tag{6.2}$$

over the family of p-distributions which satisfies the restraints

$$\Sigma\ T_1(\omega)p(\omega) = \theta_1^*$$
(6.3)
$$\Sigma\ T_2(\omega)p(\omega) = \theta_2^*$$
$$\Sigma\ T_3(\omega)p(\omega) = \theta_3^*$$

where θ_1^* , θ_2^* , θ_3^* are specified values, and $\pi(\omega)$ is a fixed distribution.

If $\pi(\omega)$ satisfies the restraints (6.3), then of course the minimum value of $I(p:\pi)$ is zero and the minimizing distribution is $p^*(\omega) = \pi(\omega)$. More generally, the minimum discrimination information theorem states that the minimizing distribution is given by

(6.4)
$$p^*(\omega) = \frac{\exp\ (\tau_1 T_1(\omega) + \tau_2 T_2(\omega) + \tau_3 T_3(\omega))\pi(\omega)}{M(\tau_1,\tau_2,\tau_3)}$$

where

(6.5)
$$M(\tau_1,\tau_2,\tau_3) = \Sigma\ \exp\ (\tau_1 T_1(\omega) + \tau_2 T_2(\omega) + \tau_3 T_3(\omega))\pi(\omega)$$

is a normalizing factor so that $\Sigma\ p^*(\omega) = 1$, and the τ's are parameters which technically are in essence undetermined Lagrange multipliers whose values are defined in terms of θ_1^* , θ_2^* , θ_3^* by

$$\theta_1^* = \frac{\partial}{\partial\tau_1}\ \ell n\ M(\tau_1,\tau_2,\tau_3)$$
$$= (\Sigma\ \exp\ (\tau_1 T_1(\omega) + \tau_2 T_2(\omega) + \tau_3 T_3(\omega))T_1(\omega)\pi(\omega))/M(\tau_1,\tau_2,\tau_3)$$
$$= \Sigma\ T_1(\omega)p^*(\omega)$$

$$\theta_2^* = \frac{\partial}{\partial\tau_2}\ \ell n\ M(\tau_1,\tau_2,\tau_3)$$
$$= (\Sigma\ \exp\ (\tau_1 T_1(\omega) + \tau_2 T_2(\omega) + \tau_3 T_3(\omega))T_2(\omega)\pi(\omega))/M(\tau_1,\tau_2,\tau_3)$$
(6.6)
$$= \Sigma\ T_2(\omega)p^*(\omega)$$

$$\theta_3^* = \frac{\partial}{\partial\tau_3}\ \ell n\ M(\tau_1,\tau_2,\tau_3)$$
$$= (\Sigma\ \exp(\tau_1 T_1(\omega) + \tau_2 T_2(\omega) + \tau_3 T_3(\omega))T_3(\omega)\pi(\omega))/M(\tau_1,\tau_2,\tau_3)$$
$$= \Sigma\ T_3(\omega)p^*(\omega)\ .$$

We can now state a number of consequences of the preceding.

We note first that $p^*(\omega)$ is a member of an exponential family of distributions generated by $\pi(\omega)$ and as such has the desirable statistical properties of members of an exponential family which include all the common and classic distributions. We may also write (6.4) as

$$\ell n \frac{p^*(\omega)}{\pi(\omega)} = -\ell n\ M(\tau_1,\tau_2,\tau_3) + \tau_1 T_1(\omega) + \tau_2 T_2(\omega) + \tau_3 T_3(\omega) \tag{6.7}$$

$$= L + \tau_1 T_1(\omega) + \tau_2 T_2(\omega) + \tau_3 T_3(\omega)$$

with $L = -\ell n\ M(\tau_1,\tau_2,\tau_3)$. The regression or log-linear expression in (6.7) for $\ell n\ (p^*(\omega)/\pi(\omega))$ with $T_1(\omega)$, $T_2(\omega)$, $T_3(\omega)$ as the explanatory variables and τ_1 , τ_2 , τ_3 as the regression coefficients plays an important role in the analysis we shall consider.

We note next that the minimum value of the discrimination information (6.2) is

$$I(p^*:\pi) = \tau_1\theta_1^* + \tau_2\theta_2^* + \tau_3\theta_3^* - \ell n\ M(\tau_1,\tau_2,\tau_3) \tag{6.8}$$

where the θ^*'s are defined in (6.3) and the τ's are determined to satisfy (6.6). Using the value in (6.7) it may be shown that if $p(\omega)$ is any member of the family of distributions satisfying (6.3), then

$$I(p:\pi) = I(p^*:\pi) + I(p:p^*) \ . \tag{6.9}$$

The pythagorean type property (6.9) plays an important role in the analysis of information tables.

7. Computational Procedures

An "experiment" has been designed and observations made resulting in a multi-dimensional contingency table with the desired classifications and categories. All the information the analyst hopes to obtain from the "experiment" is contained in the contingency table. In the process of analysis, the aim is to fit the observed table by a minimal or parsimonious number of parameters depending on some or all of the marginals, that is,

to find out how much of this total information is contained in a summary consisting of sets of marginals. Indeed, the relationship between the concept of independence or association and interaction in contingency tables and the role the marginals play is evidenced in the historical developments in the extensive literature on the analysis of contingency tables. Thus, the θ^*'s in the preceding discussion will be the marginals of interest. See Ku et al (1971).

7.1. The $T(\omega)$ Functions. The $T(\omega)$ functions for the RxSxTxU table turn out to be a basic set of simple functions and their various products. Thus, for example, the $T(\omega)$ function associated with the one-way marginal $p(2\ldots)$ is

(7.1) $$T_2^R(ijk\ell) = 1 \quad \text{for } i = 2 \text{ , any } j,k,\ell$$
$$= 0 \quad \text{otherwise}$$

since

(7.2) $$\Sigma\, p(ijk\ell)\, T_2^R(ijk\ell) = p(2\ldots) \ .$$

Similarly the $T(\omega)$ function associated with the one-way marginal $p(..3.)$, for example, is

(7.3) $$T_3^T(ijk\ell) = 1 \quad \text{for } k = 3 \text{ , any } i,j,\ell$$
$$= 0 \quad \text{otherwise}$$

since

(7.4) $$\Sigma\, p(ijk\ell)\, T_3^T(ijk\ell) = p(..3.) \ .$$

Thus for the rxsxtxu table there are

(7.5)
- $(r-1)$ linearly independent functions $T_\alpha^R(ijk\ell)$, $\alpha = 1,\ldots,r-1$
- $(s-1)$ linearly independent functions $T_\beta^S(ijk\ell)$, $\beta = 1,\ldots,s-1$
- $(t-1)$ linearly independent functions $T_\gamma^T(ijk\ell)$, $\gamma = 1,\ldots,t-1$
- $(u-1)$ linearly independent functions $T_\delta^U(ijk\ell)$, $\delta = 1,\ldots,u-1$,

since, for example,

$$\sum_{\alpha=1}^{r} \Sigma\, T_\alpha^R(ijk\ell) = rstu \ .$$

We have arbitrarily excluded the functions corresponding to $\alpha = r$, $\beta = s$, $\gamma = t$, $\delta = u$ as a matter of convenience. We could have selected $\alpha = 1$, $\beta = 1$, $\gamma = 1$, $\delta = 1$ or any other set of values.

The $T(\omega)$ function associated with the two-way marginal $p(12..)$ say, is $T_1^R(ijk\ell)\ T_2^S(ijk\ell)$ since from the definition of $T_1^R(ijk\ell)$ and $T_2^S(ijk\ell)$ it may be seen that

(7.6) $$T_1^R(ijk\ell)\ T_2^S(ijk\ell) = 1 \quad \text{for } i = 1,\ j = 2, \text{ any } k,\ell$$
$$= 0 \quad \text{otherwise}$$

and

(7.7) $$\Sigma\ p(ijk\ell)\ T_1^R(ijk\ell)\ T_2^S(ijk\ell) = p(12..)\ .$$

For convenience we shall write $T_\alpha^R(ijk\ell)\ T_\beta^S(ijk\ell) = T_{\alpha\beta}^{RS}(ijk\ell)$, etc. Thus the $T(\omega)$ function associated with any two-way marginal is a product of two appropriate functions of the set (7.5).

Similarly the $T(\omega)$ function associated with any three-way marginal will be a product of three of the appropriate functions of the set (7.5), for example,

(7.8) $$\Sigma\ p(ijk\ell)\ T_2^R(ijk\ell)\ T_1^T(ijk\ell)\ T_3^U(ijk\ell) = p(2.13)\ .$$

For convenience we shall write $T_\alpha^R(ijk\ell)\ T_\beta^S(ijk\ell)\ T_\gamma^T(ijk\ell) = T_{\alpha\beta\gamma}^{RST}(ijk\ell)$, etc.

Similarly the $T(\omega)$ function associated with any four-way marginal will be a product of four of the appropriate functions of the set (7.5), for example,

(7.9) $$\Sigma\ p(ijk\ell)\ T_2^R(ijk\ell)\ T_1^S(ijk\ell)\ T_1^T(ijk\ell)\ T_2^U(ijk\ell) = p(2112)\ .$$

For convenience we shall write $T_\alpha^R(ijk\ell)\ T_\beta^S(ijk\ell)\ T_\gamma^T(ijk\ell)\ T_\delta^U(ijk\ell) = T_{\alpha\beta\gamma\delta}^{RSTU}(ijk\ell)$.

We note that there are a total of

$$(7.10) \quad \begin{cases} N_1 = (r-1) + (s-1) + (t-1) + (u-1) \\ N_2 = (r-1)(s-1) + (r-1)(t-1) + (r-1)(u-1) + (s-1)(t-1) \\ \qquad + (s-1)(u-1) + (t-1)(u-1) \\ N_3 = (r-1)(s-1)(t-1) + (r-1)(s-1)(u-1) + (r-1)(t-1)(u-1) \\ \qquad + (s-1)(t-1)(u-1) \\ N_4 = (r-1)(s-1)(t-1)(u-1) \ , \end{cases}$$

respectively, of the simple linearly independent functions and their products two, three, four at a time. It may be verified that

$$(7.11) \qquad rstu - 1 = N = N_1 + N_2 + N_3 + N_4 \ .$$

These values of the number of $T(\omega)$ functions (or associated tau parameters) appear as appropriate degrees of freedom in the analysis of information tables.

7.2. The Estimated $p^*(\omega)$ Values. In the usual least squares regression analysis procedure, one first computes the regression coefficients and then gets the values of the estimates. In the methodology we use we reverse the procedure. Instead of trying to obtain the values of the τ's from (6.6) (which is possible) we shall first obtain the values of the estimates $p^*(\omega)$ by a straightforward convergent iterative procedure and then derive the values of the τ's from (6.7). We shall not discuss the details of the iteration here, as they are in the computer program and have been described elsewhere. The iteration may be described as successively cycling through adjustments of the marginals of interest starting with the $\pi(\omega)$ distribution until a desired accuracy of agreement between the set of observed marginals of interest and the computed marginals has been attained. See Ku et al (1971).

7.3. The τ Values or Interaction Parameters. From the definitions of the $T(\omega)$ functions in Section 7.1 it is clear that they take on only the values 0 or 1 for each value of ω . From the nature of the $T(\omega)$

functions the set of regression or log-linear equations (6.7) will have some with a single τ value which can be determined. Then there will be a set with one additional unknown value and some of the τ's already determined. These new unknown τ values can be then determined. This process of successive evaluation is carried on until all the values of τ are determined. They are also available as output of a general computer program.

8. Graphic Representation

A useful graphic representation of the log-linear regression (6.7) is given in Figure 8.1 for a 2x2x2x2 contingency table. This is the analogue of the design matrix in normal regression theory. The blank spaces in Figure 8.1 represent zero values. The $(ijk\ell)$-columns are the cell identifications in the same lexographic order as the cell entries for the estimates in the computer output. Column 1 corresponds to L which is essentially a normalizing factor. Each of the columns 2 to 16 represents the corresponding values of the $T(\omega)$ functions, columns 2 to 5 those for the one-way marginals, columns 6 to 11 those for the two-way marginals, columns 12 to 15 those for the three-way marginals, and column 16 that for the four-way marginal. For convenience the columns are also arranged in lexographic order. The tau parameter associated with the $T(\omega)$ function is given at the head of the column. The full representation with all the columns of Figure 8.1 generates the observed values. Thus the rows represent

$$(8.1)\qquad \ln \frac{p(ijk\ell)}{\pi(ijk\ell)} = \ln \frac{x(ijk\ell)}{n\pi(ijk\ell)} = L + \tau_1^i T_1^i(ijk\ell) + \ldots + \tau_{11}^{ij} T_{11}^{ij}(ijk\ell) + \ldots + \tau_{111}^{ijk} T_{111}^{ijk}(ijk\ell) + \ldots + \tau_{1111}^{ijk\ell} T_{1111}^{ijk\ell}(ijk\ell)$$

where $\pi(ijk\ell)$ in the 2x2x2x2 case is 1/2x2x2x2 and the numerical values of L and the taus depend on the observed values $x(ijk\ell)$. The design matrix corresponding to an estimate uses only those columns associated with the marginals explicit and implied in the fitting process. This is a reflection of the fact that higher order marginals imply certain

ω	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
i j k ℓ	L	τ_1^i	τ_1^j	τ_1^k	τ_1^ℓ	τ_{11}^{ij}	τ_{11}^{ik}	$\tau_{11}^{i\ell}$	τ_{11}^{jk}	$\tau_{11}^{j\ell}$	$\tau_{11}^{k\ell}$	τ_{111}^{ijk}	$\tau_{111}^{ij\ell}$	$\tau_{111}^{ik\ell}$	$\tau_{111}^{jk\ell}$	$\tau_{1111}^{ijk\ell}$
1 1 1 1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1 1 1 2	1	1	1	1		1	1		1			1				
1 1 2 1	1	1	1		1	1		1		1			1			
1 1 2 2	1	1	1			1										
1 2 1 1	1	1		1	1		1	1			1			1		
1 2 1 2	1	1		1			1									
1 2 2 1	1	1			1			1								
1 2 2 2	1	1														
2 1 1 1	1		1	1	1				1	1	1				1	
2 1 1 2	1		1	1					1							
2 1 2 1	1		1		1					1						
2 1 2 2	1		1													
2 2 1 1	1			1	1						1					
2 2 1 2	1			1												
2 2 2 1	1				1											
2 2 2 2	1															

Figure 8.1. Graphic representation.

lower order marginals, for example, the two-way marginal $x(ij..)$ implies, by summation over i and j , the one-way marginals $x(.j..)$, $x(i...)$, and the total $n = x(....)$. Thus the estimate based on fitting the one-way marginals will use only columns 1-5. The values of L and the taus for this estimate will be different from those for $x(ijk\ell)$ and depend on the estimate $x_1^*(ijk\ell)$. Thus if we denote the estimate based on fitting the one-way marginals as $x_1^*(ijk\ell)$, the representation in Figure 8.1 implies

$$
(8.2)\qquad \begin{cases} \ln \dfrac{x_1^*(1111)}{n\pi} = L + \tau_1^i + \tau_1^j + \tau_1^k + \tau_1^\ell \\ \ln \dfrac{x_1^*(1112)}{n\pi} = L + \tau_1^i + \tau_1^j + \tau_1^k \\ \quad\vdots \qquad\qquad \vdots \qquad\quad \vdots \\ \ln \dfrac{x_1^*(2222)}{n\pi} = L\ . \end{cases}
$$

the estimate based on fitting the two-way marginals will be of the form [illegible] Since the two-way marginals also imply the [illegible] values [illegible] and the taus for this estimate will be different from those for the observed values or other estimates and depend on the values of the estimate which we denote by $x_*^*(ijk\ell)$. For the estimate fitting the two-way [illegible] the representation in figure 8.1 implies

$$
(8.3)\quad
\begin{cases}
\ln \dfrac{x_*^*(1111)}{n\theta} = L + \tau_1^i + \tau_1^j + \tau_1^k + \tau_1^\ell + \tau_{11}^{ij} + \tau_{11}^{ik} + \tau_{11}^{i\ell} + \tau_{11}^{jk} + \tau_{11}^{j\ell} + \tau_{11}^{k\ell} \\
\ln \dfrac{x_*^*(1112)}{n\theta} = L + \tau_1^i + \tau_1^j + \tau_1^k + \tau_{11}^{ij} + \tau_{11}^{ik} + \tau_{11}^{jk} \\
\quad\vdots \qquad\qquad \vdots \qquad\qquad \vdots \\
\ln \dfrac{x_*^*(2222)}{n\theta} = L .
\end{cases}
$$

The representation for the uniform distribution corresponds to column 1 only.

Note that in accordance with (7.10) and (7.11)

$$
\begin{aligned}
N_1 &= 1 + 1 + 1 + 1 && = 4 \text{ (columns 2 to 5)} \\
N_2 &= 1 + 1 + 1 + 1 + 1 + 1 && = 6 \text{ (columns 6 to 11)} \\
N_3 &= 1 + 1 + 1 + 1 && = 4 \text{ (columns 12 to 15)} \\
N_4 &= 1 && = 1 \text{ (column 16)} \\
N &= 16 - 1 = 15 && = 4 + 6 + 4 + 1 .
\end{aligned}
$$

9. Analysis of Information

Although the preceding discussion has at times been in terms of probabilities, estimated probabilities or relative frequencies, in practice it has been found more convenient not to divide everything by n, the total number of occurrences, and deal with observed or estimated occurrences, that is, with $n\pi(ijk\ell) = n/rstu$, $x(ijk\ell)$, $x(i\ldots)$, $x(.j..)$, $x^*(ijk\ell) = np^*(ijk\ell)$, etc. The analysis of information is based on the fundamental relation (6.9) for the minimum discrimination information statistics. Specifically if $np_a^*(\omega) = x_a^*(\omega)$ is the minimum discrimination information estimate corresponding to a set H_a of given marginals and $x_b^*(\omega)$ is the

Copy available to DDC does not
permit fully legible reproduction

minimum discrimination information estimate corresponding to a set H_b of given marginals, where H_a is explicitly or implicitly contained in H_b, then the basic relations are

$$
\begin{aligned}
2I(x:n\pi) &= 2I(x_a^*:n\pi) + 2I(x:x_a^*) \\
2I(x:n\pi) &= 2I(x_b^*:n\pi) + 2I(x:x_b^*) \\
2I(x_b^*:n\pi) &= 2I(x_a^*:n\pi) + 2I(x_b^*:x_a^*) \\
2I(x:x_a^*) &= 2I(x_b^*:x_a^*) + 2I(x:x_b^*)
\end{aligned}
\tag{9.1}
$$

with a corresponding additive relation for the associated degrees of freedom.

In terms of the representation in (6.4) or (6.7) or Figure 8.1 as an exponential family, for our discussion, the two extreme cases are the uniform distribution for which all τ's are zero, and the observed contingency table or distribution for which all $N = rstu - 1$ τ's are needed.

Measures of the form $2I(x:x_a^*)$, that is, the comparison of an observed contingency table with an estimated contingency table, are called measures of interaction or goodness-of-fit. Measures of the form $2I(x_b^*:x_a^*)$, comparing two estimated contingency tables, are called measures of effect, that is the effect of the marginals in the set H_b but not in the set H_a or the taus in x_b^* but not in x_a^*. We note that $2I(x:x_a^*)$ tests a null hypothesis that the values of the τ parameters in the representation of the observed contingency table $x(\omega)$ but not in the representation of the estimated table $x_a^*(\omega)$ are zero and the number of these taus is the number of degrees of freedom. Similarly $2I(x_b^*:x_a^*)$ tests a null hypothesis that the values of the set of τ parameters in the representation of the estimated table $x_b^*(\omega)$ but not in the representation of the estimated table $x_a^*(\omega)$ are zero and the number of these taus is the number of degrees of freedom.

We summarize the additive relationships of the m.d.i. statistics and the associated degrees of freedom in the Analysis of Information Table 9.1.

TABLE 9.1

ANALYSIS OF INFORMATION TABLE

Component due to	Information	D.F.
H_a : Interaction	$2I(x:x_a^*)$	N_a
H_b : Effect	$2I(x_b^*:x_a^*)$	$N_a - N_b$
Interaction	$2I(x:x_b^*)$	N_b

Since measures of the form $2I(x:x_a^*)$ may also be interpreted as measures of the "variation unexplained" by the estimate x_a^*, the additive relationship leads to the interpretation of the ratio

$$\frac{2I(x:x_a^*) - 2I(x:x_b^*)}{2I(x:x_a^*)} = \frac{2I(x_b^*:x_a^*)}{2I(x:x_a^*)} \tag{9.2}$$

as the percentage of the unexplained variation due to x_a^* accounted for by the additional constraints defining x_b^*. The ratio (9.2) is thus similar to the squared correlation coefficients associated with normal distributions.

We remark that the marginals explicit and implicit of the estimated table $x_a^*(\omega)$ which form the set of restraints H_a used to generate $x_a^*(\omega)$ are the same as the corresponding marginals of the observed $x(\omega)$ table and all lower order implied marginals. It may be shown that $2I(x:x_a^*)$ is approximately a quadratic in the differences between the remaining marginals of the $x(\omega)$ table and the corresponding ones as calculated from the $x_a^*(\omega)$ table.

Similarly $2I(x_b^*:x_a^*)$ is also approximately a quadratic in the differences between those additional marginal restraints in H_b but not in H_a and the corresponding marginal values as computed from the $x_a^*(\omega)$ table.

As may be seen, because of the nature of the $T(\omega)$ functions described in Section 7.1 or indicated in Figure 8.1, the τ's are determined from the log-linear regression Equations (6.7) (see (8.2) and (10.3))

as sums and differences of values of $\ln x^*(ij|k)$. A variety of statistics have been presented in the literature for the analysis of contingency tables which are quadratics in differences of marginal values or quadratics in the τ's or the linear combinations of logarithms of the observed or estimated values. The principle of minimum discrimination information estimation and its procedures thus provides a unifying relationship since such statistics may be seen as quadratic approximations of the minimum discrimination information statistic. We remark that the corresponding approximate X^2's are not generally additive.

We mention the approximations in terms of quadratic forms in the marginals or the τ's as a possible bridge connecting the familiar procedures of classical regression analysis and the procedures proposed here to assist in understanding and interpreting the analysis of information tables. The covariance matrix of the $T(\omega)$ functions or the taus can be obtained for either the observed table or any of the estimated tables, as well as the inverse matrices as part of the output of the general computer program. See (10.4) to (10.9).

10. The 2x2 Table

It may be useful to reexamine the 2x2 table from the point of view of the preceding discussion. The algebraic details are simple in this case and exhibit the unification of the information theoretic development.

Suppose we have the observed 2x2 table in Figure 10.1

x(11)	x(12)	x(1.)
x(21)	x(22)	x(2.)
x(.1)	x(.2)	n

Figure 10.1

If we obtain the m.d.i. estimate fitting the one-way marginals, the generalized independence hypothesis is the classical independence hypothesis and the minimum discrimination information estimate is $x^*(ij) = x(i.)x(.j)/n$. The representation of the log-linear regression (6.7) as in Figure 8.1 for the full model is given in Figure 10.2. The entries in the columns τ_1 , τ_2 , τ_3

i	j	L	τ_1	τ_2	τ_3
1	1	1	1	1	1
1	2	1	1		
2	1	1		1	
2	2	1			

Figure 10.2

are, respectively, the values of the functions $T_1(ij)$, $T_2(ij)$, $T_3(ij)$ associated with the marginals $\theta_1 = x(1.)$, $\theta_2 = x(.1)$, $\theta_3 = x(11)$, and the column headed L corresponds to the normalizing factor (the negative of the logarithm of the moment-generating function as in (6.7)).

We recall the interpretation of Figure 10.2 as the log-linear relations

$$(10.1)\quad \begin{cases} \ln \dfrac{x(11)}{n\pi} = L + \tau_1 + \tau_2 + \tau_3 \\ \ln \dfrac{x(12)}{n\pi} = L + \tau_1 \\ \ln \dfrac{x(21)}{n\pi} = L + \tau_2 \\ \ln \dfrac{x(22)}{n\pi} = L \ . \end{cases}$$

From (10.1) we find

$$(10.2)\quad \begin{aligned} L &= \ln\left(x(22)/n/4\right) , \\ \tau_1 &= \ln\left(x(12)/x(22)\right) , \\ \tau_2 &= \ln\left(x(21)/x(22)\right) , \\ \tau_3 &= \ln\left(x(11)x(22)/x(12)x(21)\right) \end{aligned}$$

or

$$\tau_1 = \ell n\, x(12) - \ell n\, x(22) \; ,$$

(10.3) $$\tau_2 = \ell n\, x(21) - \ell n\, x(22) \; ,$$

$$\tau_3 = \ell n\, x(11) + \ell n\, x(22) - \ell n\, x(12) - \ell n\, x(21) \; .$$

If we call $\underline{T}$ the matrix with columns the columns of the design matrix of Figure 10.2 that is,

(10.4) $$\underline{T} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} ,$$

and define a diagonal matrix $\underline{D}$ with main diagonal the elements $x(ij)$, that is,

(10.5) $$\underline{D} = \begin{pmatrix} x(11) & 0 & 0 & 0 \\ 0 & x(12) & 0 & 0 \\ 0 & 0 & x(21) & 0 \\ 0 & 0 & 0 & x(22) \end{pmatrix} ,$$

then the estimate of the covariance matrix of $\theta_1 = x(1.)$, $\theta_2 = x(.1)$, $\theta_3 = x(11)$ for the observed contingency table is $\underline{\Sigma} = \underline{A}_{22.1}$ where

(10.6) $$\underline{A} = \begin{pmatrix} \underline{A}_{11} & \underline{A}_{12} \\ \underline{A}_{21} & \underline{A}_{22} \end{pmatrix} = \underline{T}'\underline{D}\underline{T}$$

(10.7) $$\underline{A}_{22.1} = \underline{A}_{22} - \underline{A}_{21}\,\underline{A}_{11}^{-1}\,\underline{A}_{12}$$

and $\underline{A}_{11}$ is 1×1 , $\underline{A}_{22}$ is 3×3 , $\underline{A}_{21}' = \underline{A}_{12}$ is 1×3 . It is found that

(10.8) $$\underline{\Sigma} = \begin{pmatrix} \dfrac{x(1.)x(2.)}{n} & x(11) - \dfrac{x(1.)x(.1)}{n} & \dfrac{x(11)x(2.)}{n} \\ x(11) - \dfrac{x(1.)x(.1)}{n} & \dfrac{x(.1)x(.2)}{n} & \dfrac{x(11)x(.2)}{n} \\ \dfrac{x(11)x(2.)}{n} & \dfrac{x(11)x(.2)}{n} & x(11) - \dfrac{x^2(11)}{n} \end{pmatrix} ,$$

and the inverse matrix is

$$(10.9)\quad \underline{\Sigma}^{-1} = \begin{pmatrix} \frac{1}{x(12)} + \frac{1}{x(22)} & \frac{1}{x(22)} & -\frac{1}{x(12)} - \frac{1}{x(22)} \\ \frac{1}{x(22)} & \frac{1}{x(21)} + \frac{1}{x(22)} & -\frac{1}{x(21)} - \frac{1}{x(22)} \\ -\frac{1}{x(12)} - \frac{1}{x(22)} & -\frac{1}{x(21)} - \frac{1}{x(22)} & \frac{1}{x(11)} + \frac{1}{x(12)} + \frac{1}{x(21)} + \frac{1}{x(22)} \end{pmatrix} .$$

We remark that the matrix in (10.9) is the covariance matrix of the τ's in (10.3). Similar results hold in general and for estimated tables.

Note that the value of the logarithm of the cross-product ratio, a measure of association or interaction, appears in the course of the analysis as the value of τ_3 for the observed values $x(ij)$, and that $\tau_3 = 0$ for $x^*(ij)$, the estimate under the hypothesis of independence, for which the representation as in Figure 10.2 does not involve the last column since it is obtained by fitting the one-way marginals.

The log-linear relations for the estimate $x^*(ij)$ are

$$(10.10)\qquad \begin{cases} \ell n \dfrac{x^*(11)}{n\pi} = L + \tau_1 + \tau_2 \\ \ell n \dfrac{x^*(12)}{n\pi} = L + \tau_1 \\ \ell n \dfrac{x^*(21)}{n\pi} = L + \tau_2 \\ \ell n \dfrac{x^*(22)}{n\pi} = L \quad , \end{cases}$$

where the numerical values of L, τ_1, τ_2 in (10.10) depend on x^* and differ from the values in (10.1).

The minimum discrimination information statistic to test the null hypothesis or model of independence is $2I(x:x^*)$ with one degree of freedom. In this case the quadratic approximation is

$$(10.11)\quad 2I(x:x^*) \approx \left(x(11) - \frac{x(1.)x(.1)}{n}\right)^2 \left(\frac{1}{x^*(11)} + \frac{1}{x^*(12)} + \frac{1}{x^*(21)} + \frac{1}{x^*(22)}\right) .$$

Remembering that $x^*(ij) = x(i.)x(.j)/n$, the right-hand side of (10.11) may also be shown to be

$$(10.12) \qquad \chi^2 = \Sigma\, (x(ij) - x(i.)x(.j)/n)^2 \Big/ \frac{x(i.)x(.j)}{n} ,$$

the classical χ^2-test for independence with one degree of freedom. Another test which has been proposed for the null hypothesis of no association or no interaction in the 2x2 table is

$$(10.13) \quad (\ln x(11) + \ln x(22) - \ln x(12) - \ln x(21))^2 \left(\frac{1}{x(11)} + \frac{1}{x(12)} + \frac{1}{x(21)} + \frac{1}{x(22)}\right)^{-1} ,$$

which may be shown to be a quadratic approximation for $2I(x:x^*)$ in terms of τ_3 with the covariance matrix estimated using the observed values and not the estimated values. We remark that if the observed values are used to estimate the covariance matrix then instead of the classical χ^2-test in (10.12) there is derived the modified Neyman chi-square

$$(10.14) \qquad \chi_1^2 = \Sigma\, (x(ij) - x(i.)x(.j)/n)^2 / x(ij) .$$

11. An Analysis

In order to coordinate and relate the various definitions, concepts, parameters, computational features, etc. discussed in the preceding sections we shall consider in detail the analysis of a specific contingency table.

Table 11.1 is a four-way contingency table of 14,053 men in a training program , cross-classified on the variables home region , level of education, race and program completion. We denote the occurrences in the four-way cross-classification or contingency Table 11.1 by $x(ijk\ell)$ with the notation

Variable	Index	1	2	3	4
Home Region	i	East	North	West	South
Level of Education	j	Below H.S.	H.S.	Above H.S.	
Race	k	White	Non-white		
Program	ℓ	Failed	Passed		

TABLE II.1

PROGRAM COMPLETION

x(ijkℓ)

i		East						North					
j		Below H.S.		H.S.		Above H.S.		Below H.S.		H.S.		Above H.S.	
k		W	Non-w	W	Non-w	W	Non-w	W	Non-w	W	Non-w	W	Non-w
ℓ	F	62	10	44	4	8	0	20	4	19	2	[illegible]	[illegible]
	P	944	133	1881	195	320	12	870	103	2205	148	471	2[illegible]

i		West						South					
j		Below H.S.		H.S.		Above H.S.		Below H.S.		H.S.		Above H.S.	
k		W	Non-w	W	Non-w	W	Non-w	W	Non-w	W	Non-w	W	Non-w
ℓ	F	14	2	9	0	0	1	42	9	34	16	[illegible]	[illegible]
	P	567	40	1350	94	421	19	952	228	1741	506	46[illegible]	[illegible]

For this data we are interested in the possible relationship of success in training as a dependent variable on the independent or explanatory variables home region, level of education, and race. To obtain a smoothed estimate of the observed cross-classification utilizing significant effects and interactions we shall examine a sequence of minimum discrimination information estimates based on nested sets of fitted marginals. That is, each successive estimate uses a set of marginals which explicitly or implicitly contains the marginals of the preceding estimate and also additional ones to determine the effect of the additional marginals or their associated interaction tau parameters. The analysis of information table permits us to judge the significance or non-significance of these effects or interaction tau parameters.

11.1 Fitting Nested Sets of Marginals. Since we are interested in the possible relationship of success in training on home region, level of education and race, we first fit the marginals $x(ijk.)$, $x(...\ell)$ since the corresponding estimate $x^*(ijk\ell) = x(ijk.)x(...\ell)/n$ is that under the null hypothesis or model of independence of success and the joint variable (home region, level of education, race) or no interaction between success and the joint variable. In other words we first want to determine whether the 24 columns of Table 11.1 are homogeneous or not with respect to the underlying probabilities of passing or failing. The associated m.d.i. statistic is

$$2I(x:x^*) = 2 \sum\sum\sum\sum x(ijk\ell) \ln(x(ijk\ell)/x^*(ijk\ell)) = 160.55$$

with 23 degrees of freedom. We reject the hypothesis of independence or no interaction. We therefore shall look for explanatory effects.

In Figure 11.1 there is given the complete schematic for the log-linear representations. The representation for the estimate of joint independence $x^*(ijk\ell) = x(ijk.)x(...\ell)/n$ uses columns 1-17, 21-22, 26-31 corresponding to all the marginals explicit and implicit in the fitted marginal constraints. We can also interpret $2I(x:x^*)$ as testing a null hypothesis or model that the 23 tau parameters in the representation of x but not in x^* are zero, that is, the parameters corresponding to columns 18-20, 23-25, 32-48.

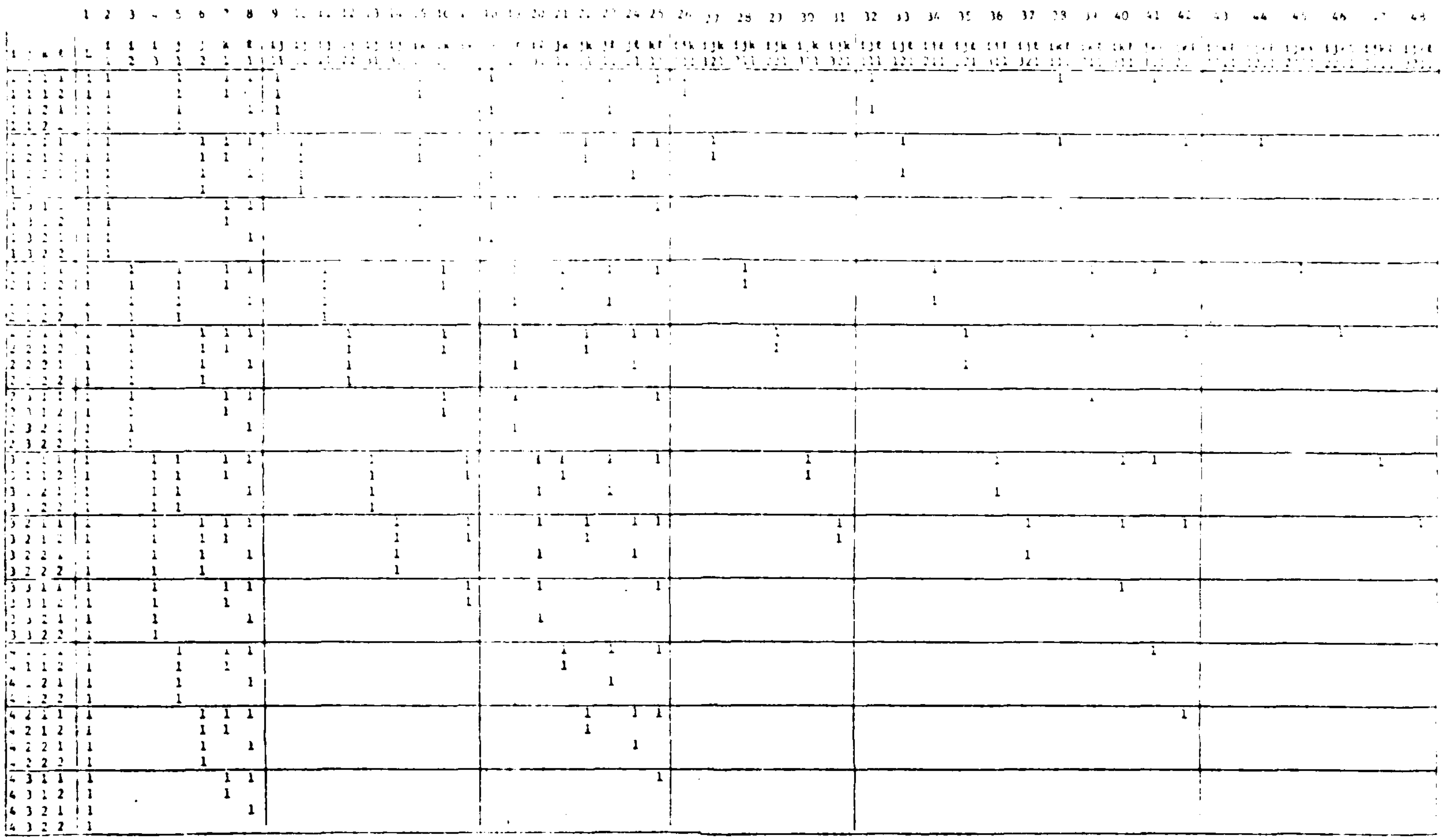

Figure II.1. Log-linear Representation.

The value of $2I(x:x^*)$ is so large that we reject the model of joint independence. We therefore proceed to fit a sequence of nested marginals all including $x(ijk.)$ and various combinations of two- and three-way marginals containing success with other variables. We summarize some results in the truncated Analysis of Information Table 11.2. We have not included all the intermediate fitting sequences for conciseness. We remark that although the measure of the effect of additional marginals or their associated parameters may vary according to the sequence in which they have been added, significant effects tend to remain significant and non-significant effects tend to stay non-significant so that the first overall survey should determine the estimates and interaction parameters which warrant further investigation. For example, the effect of adding $x(..k\ell)$ to $x(ijk.)$, $x(i..\ell)$, $x(.j.\ell)$ is given in Analysis of Information Table 11.3 as $2I(x_f^*:x_a^*) =$ 1.410 with one degree of freedom, but the effect of adding $x(..k\ell)$ to $x(ijk.)$, $x(ij.\ell)$ is given in Analysis of Information Table 11.2 as $2I(x_e^*:x_m^*) = 1.239$ with one degree of freedom. In neither case is the effect or the corresponding tau parameter $\tau_{11}^{k\ell}$ significant.

The columns of Figure 11.1 which occur in the log-linear representations of the estimates retained in Analysis of Information Table 11.2 are

Marginals Fitted	Estimate	Columns of Figure 11.1
$x(ijk.)$, $x(...\ell)$	x^*	1-17, 21-22, 26-31
$x(ijk.)$, $x(i..\ell)$, $x(.j.\ell)$	x_a^*	1-24, 26-31
$x(ijk.)$, $x(ij.\ell)$	x_m^*	1-24, 26-37
$x(ijk.)$, $x(ij.\ell)$, $x(..k\ell)$	x_e^*	1-37 .

From the analytic form of the log-linear representation or by taking differences of appropriate rows of Figure 11.1 within the columns used for the estimate, the log-odds of fail to pass for each of the estimates are given by the respective parametric representations in (11.1) where the superscripts relate to the variables and the subscripts range over the possible indices. The values of the parameters depend of course on the corresponding estimate.

TABLE 11.2

ANALYSIS OF INFORMATION TABLE

Component Due to	Information	D.F.
$x(ijk.), x(...\ell)$	$2I(x:x^*) = 160.551$	23
a) $x(ijk.), x(i..\ell), x(.j.\ell)$	$2I(x_a^*:x^*) = 138.732$	5
	$2I(x:x_a^*) = 21.819$	18
m) $x(ijk.), x(ij.\ell)$	$2I(x_m^*:x_a^*) = 7.384$	6
	$2I(x:x_m^*) = 14.435$	12
e) $x(ijk.), x(ij.\ell), x(..k\ell)$	$2I(x_e^*:x_m^*) = 1.239$	1
	$2I(x:x_e^*) = 13.196$	11

$$\frac{2I(x:x^*) - 2I(x:x_a^*)}{2I(x:x^*)} = \frac{138.732}{160.551} = 0.86$$

$$\frac{2I(x:x^*) - 2I(x:x_m^*)}{2I(x:x^*)} = \frac{146.116}{160.551} = 0.91$$

$$\frac{2I(x:x^*) - 2I(x:x_e^*)}{2I(x:x^*)} = \frac{147.355}{160.551} = 0.92$$

TABLE 11.3

ANALYSIS OF INFORMATION TABLE

Component Due to	Information	D.F.
a) $x(ijk.), x(i..\ell), x(.j.\ell)$	$2I(x:x_a^*) = 21.819$	18
f) $x(ijk.), x(i..\ell), x(.j.\ell), x(..k\ell)$	$2I(x_f^*:x_a^*) = 1.410$	1
	$2I(x:x_f^*) = 20.409$	17

$$\ln \frac{x_a^*(ijk1)}{x_a^*(ijk2)} = \tau_1^{\ell} + \tau_{i1}^{i\ell} + \tau_{j1}^{j\ell}$$

(11.1)
$$\ln \frac{x_m^*(ijk1)}{x_m^*(ijk2)} = \tau_1^{\ell} + \tau_{i1}^{i\ell} + \tau_{j1}^{j\ell} + \tau_{ij1}^{ij\ell}$$

$$\ln \frac{x_e^*(ijk1)}{x_e^*(ijk2)} = \tau_1^{\ell} + \tau_{i1}^{i\ell} + \tau_{j1}^{j\ell} + \tau_{11}^{k\ell} + \tau_{ij1}^{ij\ell}$$

We recall that parameters with indices $i = 4$ and/or $j = 3$ and/or $k = 2$ and/or $\ell = 2$ are by convention set equal to zero.

We remark that $x_m^*(ijk\ell)$, determined by fitting the marginals $x(ijk.)$, $x(ij.\ell)$, is expressible explicitly as

(11.2)
$$x_m^*(ijk\ell) = x(ijk.)x(ij.\ell)/x(ij..)$$

and is the estimate under a null hypothesis that race and success are conditionally independent given home region and level of education. In Analysis of Information Table 11.2 the value $2I(x:x_m^*) = 14.435$, 12 degrees of freedom, indicates an acceptable fit of this model. Furthermore, $2I(x_e^*:x_m^*) = 1.239$, one degree of freedom, implies that the additional effect of the marginal $x(..k\ell)$ is not significant or that in the parametric representation of the log-odds in (11.1) the parameter $\tau_{11}^{k\ell}$ measuring the effect of race on the dependent variable success is not significant. We therefore investigate the estimate x_m^* in greater detail. The values of $x_m^*(ijk\ell)$ are given in Table 11.4.

In the expression for the log-odds under x_m^* in(11.1) τ_1^{ℓ} is an overall average, $\tau_{i1}^{i\ell}$ and $\tau_{j1}^{j\ell}$ are the effects of home region and level of education on program completion and $\tau_{ij1}^{ij\ell}$ is the interaction effect of home region x level of education on program completion. The numerical values of the tau parameters are given in Table 11.5. We recall that by convention parameters with an index corresponding to $i = 4$ and/or $j = 3$ and/or $\ell = 2$ are equal to zero.

TABLE 11.4

PROGRAM COMPLETION

$x^*_m(ijk\ell)$

	i	East						North					
	j	Below H.S.		H.S.		Above H.S.		Below H.S.		H.S.		Above H.S.	
	k	W	Non-w	W	Non-w	W	Non-w	W	Non-w	W	Non-w	W	Non-w
ℓ	F	63.039	8.961	43.503	4.497	7.718	0.282	21.424	2.576	18.736	1.264	5.713	0.287
	P	942.960	134.039	1881.497	194.503	320.282	11.718	868.575	104.424	2204.264	148.736	471.237	23.713

	i	West						South					
	j	Below H.S.		H.S.		Above H.S.		Below H.S.		H.S.		Above H.S.	
	k	W	Non-w	W	Non-w	W	Non-w	W	Non-w	W	Non-w	W	Non-w
ℓ	F	14.921	1.079	8.418	0.582	0.955	0.045	41.181	9.819	38.604	11.396	5.368	0.632
	P	566.078	40.921	1350.582	93.418	420.045	19.955	952.819	227.181	1736.396	512.604	461.632	54.368

TABLE 11.5

VALUES OF PARAMETERS IN LOG-ODDS FOR x_m^* IN (11.1)

$\tau_{1}^{\ell} = -4.454347$	$\tau_{111}^{ij\ell} = -0.292478$
$\tau_{11}^{i\ell} = 0.728653$	$\tau_{121}^{ij\ell} = -0.689433$
$\tau_{21}^{i\ell} = 0.041549$	$\tau_{211}^{ij\ell} = -0.602435$
$\tau_{31}^{i\ell} = -1.632427$	$\tau_{221}^{ij\ell} = -1.003045$
$\tau_{11}^{j\ell} = 1.312903$	$\tau_{311}^{ij\ell} = 1.137932$
$\tau_{21}^{j\ell} = 0.648130$	$\tau_{321}^{ij\ell} = 0.360697$

From the parametric representation of the log-odds in (11.1) and the values in Table 11.5 one can determine differences in the log-odds associated with changes in various categories. Thus the differences in the log-odds (fail to pass) as one changes the home region, for fixed level of education, are given by

	E-N	E-W	E-S
Below H.S.	0.9970	0.7287	0.4362
H.S.	1.0007	1.3110	0.0392
Above H.S.	0.6871	2.3611	0.7287

The differences in the log-odds as one changes the level of education for fixed home region are given by

	Below H.S.-H.S.	H.S.-Above H.S.
East	1.0617	-0.0413
North	1.0654	-0.3549
West	1.4420	1.0088
South	0.6648	0.6481

For easier interpretation, we convert the log-odds values to ratios of the odds of failure.

	E/N	E/W	E/S
Below H.S.	2.7	2.1	1.6
H.S.	2.7	3.7	1.0
Above H.S.	2.0	10.6	2.1

	Below H.S./H.S.	H.S./Above H.S.
East	2.9	0.96
North	2.9	0.70
West	4.2	2.7
South	1.9	1.9

Note that the odds of failure in training of a man with home region East and Above H.S. level of education are 10.6 times the odds of a man with the same level of education but home region West. Men with home region East or North but with level of education H.S. do better than men with same home region but Above H.S. level of education.

We have also computed the odds of failure $x_m^*(ijk1)/x_m^*(ijk2)$ and listed the results in increasing values. The odds are expressed to 1,000, that is, 5 to 1,000, 6 to 1,000, etc.

Home region	Level of Education	Odds
West	Above H.S.	2
West	H.S.	6
North	H.S.	9
South	Above H.S.	12
North	Above H.S.	12
South	H.S.	22
East	H.S.	23
East	Above H.S.	24
North	Below H.S.	25
West	Below H.S.	26
South	Below H.S.	43
East	Below H.S.	67

Note that the overall odds of failure for this data are 311/13742 = 0.0226 or 2%.

For ease of comparison and inference, we also list the foregoing results by home region and level of education.

	West	North	South	East
Above H.S.	2	12	12	24
H.S.	6	9	22	23
Below H.S.	26	25	43	67

12. Outliers

We define outliers as observations in one or more cells of a contingency table which apparently deviate significantly from a fitted model. These outliers may lead one to reject a model which fits the other observations. For example, in multi-dimensional contingency tables in which time or age is one of the classifications there may occur an age effect such that a model may be rejected for the entire table but a model taking the possible age effect into account may lead to an acceptable partitioning of the model.

In other cases even though a model seems to fit, the outliers contribute much more than reasonable to the measure of deviation between the data and the fitted values of the model. In other words, the outliers make up a large percentage of the "unexplained variation" $2I(x:x^*)$.

A clue to possible outliers is provided by the output of the computer program. In the computer output for each estimate five entries are

listed for each cell. The fourth of these is titled OUTLIER and its numerical value provides a lower bound for the decrease in the corresponding $2I(x:x^*)$ if that cell were not included in the fitting procedure. Since the reduction in the degrees of freedom is one for each omitted cell, values of OUTLIER greater than say 3.5 are of interest. The basis for the OUTLIER computation and interpretation follows. Let x_a^* denote the minimum discrimination information estimate subject to certain marginal restraints. Let x_b^* denote the minimum discrimination information estimate subject to the same marginal restraints as x_a^* except that the value $x(\omega_1)$, say, is not included, so that $x_b^*(\omega_1) = x(\omega_1)$. The basic additivity property of the minimum discrimination information statistics states that

$$2I(x:x_a^*) = 2I(x_b^*:x_a^*) + 2I(x:x_b^*)$$

or

$$2I(x:x_a^*) - 2I(x:x_b^*) = 2I(x_b^*:x_a^*) \ .$$

These results are summarized in the Analysis of Information Table 12.1

TABLE 12.1

ANALYSIS OF INFORMATION TABLE

Component due to	Information	D.F.
H_a :	$2I(x:x_a^*)$	N_a
H_b : Same as H_a but omitting $x(\omega_1)$	$2I(x_b^*:x_a^*)$	1
	$2I(x:x_b^*)$	$N_b = N_a - 1$

But

$$2I(x_b^*:x_a^*) = 2\left(x_b^*(\omega_1) \ln \frac{x_b^*(\omega_1)}{x_a^*(\omega_1)} + \sum_{\Omega-\omega_1} x_b^*(\omega) \ln \frac{x_b^*(\omega)}{x_a^*(\omega)} \right)$$

(12.1)

$$= 2\left(x(\omega_1) \ln \frac{x(\omega_1)}{x_a^*(\omega_1)} + \sum_{\Omega-\omega_1} x_b^*(\omega) \ln \frac{x_b^*(\omega)}{x_a^*(\omega)} \right) ,$$

and using the convexity property which implies that

$$(12.2) \qquad \sum_{\Omega-\omega_1} x_b^*(\omega)\ \ln \frac{x_b^*(\omega)}{x_a^*(\omega)} \geq \left(\sum_{\Omega-\omega_1} x_b^*(\omega)\right) \ln \frac{\left(\sum_{\Omega-\omega_1} x_b^*(\omega)\right)}{\left(\sum_{\Omega-\omega_1} x_a^*(\omega)\right)}$$

$$= (n - x_b^*(\omega_1))\ \ln \frac{n - x_b^*(\omega_1)}{n - x_a^*(\omega_1)} \quad ,$$

we get from (12.1)that

$$(12.3) \qquad 2I(x_b^*:x_a^*) \geq 2\left(x(\omega_1)\ \ln \frac{x(\omega_1)}{x_a^*(\omega_1)} + \left(\sum_{\Omega-\omega_1} x_b^*(\omega)\right) \ln \frac{\left(\sum_{\Omega-\omega_1} x_b^*(\omega)\right)}{\left(\sum_{\Omega-\omega_1} x_a^*(\omega)\right)}\right)$$

$$= 2\left(x(\omega_1)\ \ln \frac{x(\omega_1)}{x_a^*(\omega_1)} + (n - x(\omega_1))\ln \frac{n - x(\omega_1)}{n - x_a^*(\omega_1)}\right).$$

The last value can be computed and is listed as the OUTLIER entry for each cell of the computer output for the estimate x_a^*. We remark that a separate outlier computation for each cell is time consuming.

The ratio

$$(12.4) \qquad \frac{2I(x:x_a^*) - 2I(x:x_b^*)}{2I(x:x_a^*)} = \frac{2I(x_b^*:x_a^*)}{2I(x:x_a^*)}$$

then indicates the percentage of the "unexplained variation" due to the outlier value.

We shall illustrate the outlier procedures and analysis using data originally given by H.F. Dorn (The relationship of cancer of the lung and the use of tobacco. American Statistician, 8(1954), 7-13) and analyzed by J. Cornfield (A statistical problem arising from retrospective studies, Proc. 3rd Berkeley Symposium 4(1956), 135-148).

Cox (1970) considers a model in which the logistic difference is the same for k independent 2x2 contingency tables. He defines a residual which should behave approximately like the residuals for a random sample from the unit normal distribution. He illustrates his graphical analysis with the data originally given by Dorn and analyzed by Cornfield as mentioned above.

In Table 12.1 are listed the observations from 14 retrospective studies on the possible association between smoking and lung cancer. We denote the occurrences in the three-way 14x2x2 contingency table by $x(ijk)$ with the notation

Variable	Index	1	2	3	...	14
Study	i	No. 1	No. 2	No. 3	...	No. 14
Patients	j	Control	Lung cancer			
Smoking	k	Nonsmoker	Smoker			

Does this data show association between smoking and lung cancer, and if so, is the association homogeneous over the 14 studies? Here the measure of association is the logarithm of the cross-product ratio.

Table 12.1

Fourteen Retrospective Studies of the Association Between Smoking And Lung Cancer

Study	Control Patients		Lung Cancer Patients	
	Non-Smokers	Smokers	Non-Smokers	Smokers
1	14	72	3	83
2	43	227	3	90
3	19	81	7	129
4	125	397	12	70
5	131	299	32	412
6	114	666	8	597
7	12	174	5	88
8	61	1296	7	1350
9	27	106	3	60
10	81	534	18	459
11	54	246	4	724
12	56	462	19	499
13	636	1729	[illegible]	451
14	28	259	5	260

The hypothesis of conditional independence given the study

(12.5) $$H_a: \frac{p(ijk)}{p(i\cdot\cdot)} = \frac{p(ij\cdot)}{p(i\cdot\cdot)} \frac{p(i\cdot k)}{p(i\cdot\cdot)}$$

imposes the restraints on the estimate $x_a^*(ijk)$ that

(12.6) $$x_a^*(ij\cdot) = x(ij\cdot) \text{ and } x_a^*(i\cdot k) = x(i\cdot k) \quad .$$

In fact $x_a^*(ijk)$ may be explicitly represented by

(12.7) $$x_a^*(ijk) = x(ij\cdot)x(i\cdot k)/x(i\cdot\cdot) \quad .$$

Similarly the hypothesis of no second-order interaction

(12.8) $$H_2: \ p(ijk) = a(ij)b(ik)c(jk)$$

imposes the restraints on the estimate $x_2^*(ijk)$ that

(12.9) $$x_2^*(ij\cdot) = x(ij\cdot), \ x_2^*(i\cdot k) = x(i\cdot k), \ x_2^*(\cdot jk) = x(\cdot jk) \quad .$$

The estimate $x_2^*(ijk)$ cannot be represented as an explicit function of the observed marginals.

In this study, the minimum discrimination information statistics are log-likelihood ratio chi-squares and the associated Analysis of Information table permits us to test the goodness-of-fit of the estimates, also the effect

of adding the marginal restraint $x(\cdot jk)$ to the marginal restraints $x(ij.)$ and $x(i.k)$, or the significance of the common interaction parameter for all 14 studies associated with $x(.jk)$

$$(12.10)\qquad \tau_{11}^{jk} = \ln \frac{x_2^*(i11)\,x_2^*(i22)}{x_2^*(i12)\,x_2^*(i21)}\ , \qquad i=1,2,\ldots,14.$$

We recall that the log-odds of control to lung cancer for the estimates x_a^* and x_2^* are given by the log-linear representations

$$\ln \frac{x_a^*(i1k)}{x_a^*(i2k)} = \tau_1^j + \tau_{i1}^{ij}$$

(12.11)

$$\ln \frac{x_2^*(i1k)}{x_2^*(i2k)} = \tau_1^j + \tau_{i1}^{ij} + \tau_{11}^{jk}$$

where the values of the tau parameters depend of course on x_a^* and x_2^*.

Table 12.2

Analysis of Information

Component due to	Information	D.F.
$H_a: x(ij.), x(i.k)$	$2I(x:x_a^*)=549.74$	14
$H_2: x(ij.), x(i.k), x(.jk)$	$2I(x_2^*:x_b^*)=49$[illegible]$.55$	1
	$2I(x:x_2^*)=5$[illegible]$.19$	13

The value of $2I(x:x_a^*)$ when compared to the α-th percentile of a χ^2_{14} distribution suggests that the null hypothesis of no association between smoking and lung cancer conditioned on the study is false. This conditional hypothesis allows the accumulation of information from different studies without imposing the requirement that the population characteristics of each study be similar. The rejection of this conditional independence hypothesis is of course expected. The degree of departure from independence is functionally dependent on the study. Is this dependence the result of a small subset of the studies which are substantially different from the remainder, or does the departure vary along a continuum?

The value of $2I(x_2^*:x_b^*)$ suggests that in accordance with (12.10) the value of $\tau_{11}^{jk}=1.687$ is significantly different from zero. Moreover, $2I(x:x_2^*)$ is also significant when compared to the α-th percentile of a χ^2_{13} distribution. The value of $2I(x:x_2^*)$ suggests that we reject the null hypothesis of no second-order interaction, that is, the model with a common value of the interaction parameter τ_{11}^{jk}, is not a good fit. The values of x_2^* are given in Table 12.3. We now proceed to determine the outliers.

Table 12.3

$x_j^2(ijk)$

Study	Control Patients: Non-Smokers	Control Patients: Smokers	Lung Cancer Patients: Non-Smokers	Lung Cancer Patients: Smokers
1	14.01	71.99	2.99	83.01
2	42.86	227.14	3.14	89.86
3	19.99	80.11	6.01	129.99
4	132.16	389.83	4.84	77.16
5	130.03	300.00	32.97	410.99
6	105.06	674.94	16.94	588.06
7	15.47	170.53	1.54	91.47
8	57.06	1299.93	10.94	1346.08
9	27.15	105.85	2.85	60.15
10	85.21	529.79	13.79	463.21
11	38.62	261.39	19.38	708.62
12	62.23	455.77	12.77	505.23
13	643.32	1721.66	31.69	458.33
14	27.84	259.16	5.17	259.84

Examination of the computer output for x_2^* using all 14 studies showed a largest OUTLIER value of 18.14 for the cell (11,2,1). A new estimate fitting the marginals $x(ij.)$, $x(i.k)$, $x(.jk)$ and omitting the cell (11,2,1) was obtained. In fact Study 11 was omitted because with the constraints for the new estimate $x_b^*(11,j,k) = x(11,j,k)$. Since this estimate yielded

$$(12.12)\quad 2I(x:x_b^*) = 28.40, \quad 12 \text{ d.f.}$$

the deletion procedure was continued. We summarize the results in Table 12.4 and Analysis of Information Table 12.5.

Table 12.4

Fitting $x(ij.)$, $x(i.k)$, $x(.jk)$ with sequential deletion of studies

	Largest OUTLIER			
Study No.s	Cell	Value	Information	D.F.
1-14	(11,2,1)	18.14	$2I(x:x_2^*)=55.19$	13
1-10,12-14	(6,2,1)	7.89	$2I(x:x_b^*)=28.40$	12
1-5,7-10,12-14	(4,2,1)	4.87	$2I(x:x_c^*)=18.03$	11
1-3,5,7-10,12-14	(7,2,1)	3.91	$2I(x:x_d^*)=11.94$	10
1-3,5,8-10,12-14			$2I(x:x_e^*)= 7.03$	9

Table 12.5

Analysis of Information

Component due to	Information	D.F.
All 14 studies	$2I(x:x_2^*)=55.19$	13
Less 11	$2I(x_b^*:x_2^*)=26.79$	1
	$2I(x:x_b^*)=28.40$	12
Less 11,6	$2I(x_c^*:x_b^*)=10.37$	1
	$2I(x:x_c^*)=18.03$	11
Less 11,6,4	$2I(x_d^*:x_c^*)= 6.08$	1
	$2I(x:x_d^*)=11.94$	10
Less 11,6,4,7	$2I(x_e^*:x_d^*)= 4.92$	1
	$2I(x:x_e^*)= 7.03$	9

Since $(2I(x:x_2^*) - 2I(x:x_e^*))/2I(x:x_2^*)$ $= 2I(x_e^*:x_2^*)/2I(x:x_2^*) = 48.16/55.19=0.87$ we see that the four studies numbered 4,6,7,11 contributed 87% of the "unexplained variation" $2I(x:x_2^*)$. The values of the estimate x_e^* are given in Table 12.6. The value of the log cross-product ratio is

$$(12.13)\quad \hat{\rho}_{11}^{jk} = \ell n \frac{x_e^*(i11)\ x_e^*(i22)}{x_e^*(i12)\ x_e^*(i21)} = 1.55,\ i=1-3,5,8-10,12-14 .$$

Table 12.6

$x_c^*(ijk)$

Study	Control Patients		Lung Cancer Patients	
	Non-Smokers	Smokers	Non-Smokers	Smokers
1	13.69	72.32	3.32	82.69
2	42.46	227.53	3.54	89.47
3	19.40	80.60	6.60	129.40
5	126.85	303.18	36.15	407.83
8	55.79	1301.21	12.21	1344.80
9	26.80	106.20	3.20	59.80
10	83.61	531.39	15.39	461.62
12	60.81	457.20	14.19	503.81
13	639.35	1725.64	35.66	454.36
14	27.24	259.76	5.76	259.24

We note that Cox (1970) in analyzing the data of Table 12.1 concluded that studies 8, 6, and 11 were outliers. For the 14 studies he found a residual sum of squares 47.7 with 13 degrees of freedom. With studies 8, 6, and 11 omitted he found a residual sum of squares 15.1 with 10 degrees of freedom. (Cox (1970) p. 83 gives the degrees of freedom as 11, a misprint).

Following the procedure described when Studies 6, 8, and 11 were omitted the results led to the Analysis of Information Table 12.7. Note that omitting Studies 11, 6, 4 as per Table 12.5 accounts for more of the unexplained variation.

Table 12.7

Analysis of Information

Component due to	Information	D.F.
All 14 studies	$2I(x:x_2^*)=55.19$	13
Less 6,8,11	$2I(x_f^*:x_2^*)=41.62$	3
	$2I(x:x_f^*)=13.57$	10

The sequential procedure discussed herein was also applied to data relating father and son professions

published by Karl Pearson (1904), "On the theory of contingency and its relation to association and normal correlation," reprinted in Karl Pearson's Early Papers, Cambridge University Press, 1948, and considered by Fienberg (1969) and Good (1956). Using the Pearson data Fienberg obtained an $X^2 = 184.9$ with 44 out of 196 cells deleted whereas the sequential procedure led to an $X^2 = 155.3$ with 25 cells deleted.

13. Zero Marginals

As may be noted from the analysis in Section 11, zero occurrences in cells of the observed contingency table present no special problem provided that no marginal entering into the fitting specification is zero. When the latter is the case, however, the interpretation may be distorted because of inflated degrees of freedom. A procedure to circumvent this problem is similar to that used for getting revised estimates when outliers are indicated. We shall present the procedure in terms of a specific example.

The following data resulted from a study of Christmas tree consumption. We are indebted to Dipl. Forstwirt Dietrich V. Staden, Institut f. Forstbenutzung, Universitaet Goettingen for the data and permission to use it. In Table 13.1 are listed responses to the question "Did you have a Christmas tree in your apartment/house last year or not?" according to size of household and size of city. We denote the occurrences

in the three-way 2x9x5 contingency table by x(ijk) with the notation

Variable	Index	1	2	3	4	5	6	7	8
Tree	i	Yes	No						
Household size	j	1	2	3	4	5	6	7	8
City size	k	<2000	2000 to 20000	20000 to 100000	100000 to 500000	500000 or more			

For a 2x9x5 RxCxD contingency table we compute an estimate under a hypothesis of no second-order interaction by fitting all the two-way marginals. Call this estimate $x_2^*(ijk)$. A test for the null hypothesis of no second-order interaction is given by

$$2\Sigma\Sigma\Sigma\; x(ijk)\; \ln \frac{x(ijk)}{x_2^*(ijk)} = 2I(x:x_2^*)\quad,\quad 32\text{ D.F.}$$

If there is no second-order interaction then the associations between R and C, R and D, C and D are the same for all values of the third variable, that is,

$$\ln \frac{x_2^*(1jk)\,x_2^*(29k)}{x_2^*(2jk)\,x_2^*(19k)} \text{ depends only on } j,$$

$$\ln \frac{x_2^*(1jk)\,x_2^*(2j5)}{x_2^*(2jk)\,x_2^*(1j5)} \text{ depends only on } k,$$

$$\ln \frac{x_2^*(ijk)\,x_2^*(i95)}{x_2^*(ij5)\,x_2^*(i9k)} \text{ is independent of } i.$$

Within this model a test whether the marginal $x(i \cdot k)$ contributes significantly is obtained by computing an estimate fitting the marginals $x(ij\cdot)$, $x(\cdot jk)$ only. Call this estimate $x_b^*(ijk)$, which can be expressed as $x_b^*(ijk) = x(ij\cdot)x(\cdot jk)/x(\cdot j\cdot)$. We recognize $x_b^*(ijk)$ as the estimate under **an** hypothesis of conditional independence of R and D given C. We now have Analysis of Information Table 13.2.

Table 13.2

Component due to	Information	D.F.
Conditional independence of R and D given C	$2I(x:x_b^*)$	36
Effect of $x(i \cdot k)$ given $x(ij\cdot)$ and $x(\cdot jk)$	$2I(x_2^*:x_b^*)$	4
No second-order interaction	$2I(x:x_2^*)$	32

For the particular data in question however, because $x(ijk)=0$ for $j=6,7,8,9$, $i=2$ and also for some of $i=1$, $j=7,8,9$, the estimates for the entries corresponding to $x^*(ijk)$ for $j=6,7,8,9$ both for x_2^* and x_b^* will not differ from the observed value. Accordingly let us compute an estimate $x_L^*(ijk)$ which is obtained by fitting the two-way marginals of the 2x5x5 table $j=1,2,3,4,5$ and $x_L^*(ijk)=x(ijk)$, $j=6,7,8,9$. Similarly let $x_c^*(ijk)=x(ij\cdot)x(\cdot jk)/x(\cdot j\cdot)$ for the 2x5x5 table $j=1,2,3,4,5$ and $x_c^*(ijk)=x(ijk)$, $j=6,7,8,9$.

We now find

Table 13.3

Component due to	Information	D.F.
Conditional independence of R and D given C	$2I(x:x_e^*)=25.532$	20
Effect of x(i·k) given x(ij·) and x(·jk)	$2I(x_f^*:x_e^*)=5.821$	4
No second-order interaction	$2I(x:x_f^*)=19.711$	16

Note the reduction in the degrees of freedom between Table 13.2 and Table 13.3. It is also interesting to note that when actually carrying out the procedures for Table 13.2, the same estimates and statistics were obtained as for Table 13.3. See Table 13.4 and 13.5, Table 13.6 and 13.7.

It seems reasonable to conclude that the purchase of a Christmas tree is independent of the size of the city given the size of the household (j=1,2,3,4,5) and households of size 6,7,8,9 seem almost sure to buy Christmas trees.

The log-odds for the purchase of a Christmas tree as a function of household size is given in Table 13.8. The probability estimate for a purchase as a function of household size is given in Table 13.9.

Table 13.8

$$\ln\ (x_e^*(1jk)/x_e^*(2jk)) = \ln\ (x(1j\cdot)/x(2j\cdot))$$

j=1	-0.2586
2	0.8662
3	2.1702
4	3.4012
5	2.3716

Table 13.9

$x(1j\cdot)/x(\cdot j\cdot)$

j=1	61/140	=	0.4357
2	214/304	=	0.7039
3	219/244	=	0.8975
4	180/186	=	0.9677
5	75/82	=	0.9146

For more complex situations there is also the log-linear analysis, which is of course available for this problem too, but it would not add anything to the analysis of this particular data.

Table 13.1
original data x(ijk)
k

i j	1	2	3	4	5	
1 1	4	7	6	12	32	[illegible]
1 2	18	37	[illegible]	41	[illegible]	214
1 3	20	45	52	41	61	210
1 4	25	40	38	32	[illegible]	180
1 5	11	31	14	11	8	7
1 6	4	8	12	1	6	31
1 7	3	0	2	1	1	7
1 8	2	0	1	0	1	4
1 9	0	1	0	0	1	2
2 1	4	13	9	13	40	79
2 2	5	18	19	15	33	90
2 3	3	8	0	4	10	25
2 4	0	3	1	0	2	6
2 5	1	1	2	2	1	7
2 6	0	0	0	0	0	0
2 7	0	0	0	0	0	0
2 8	0	0	0	0	0	0
2 9	0	0	0	0	0	0

Table 13.4
$x_2^*(ijk)$ [illegible]X5
k

1	2	3	4	5
4.066	7.835	7.815	12.046	29.240
17.253	38.835	56.204	40.865	63.844
21.123	46.332	47.951	40.955	62.588
24.364	41.268	38.056	31.098	45.213
11.193	28.680	14.974	12.036	8.115
4.000	8.000	12.000	1.000	6.000
3.000	0.000	2.000	1.000	1.000
2.000	0.000	1.000	0.000	1.000
0.000	1.000	0.000	0.000	1.000
3.933	12.167	7.184	12.953	42.763
5.747	19.166	17.795	15.134	32.157
1.877	6.618	4.050	4.046	8.410
0.636	1.731	0.945	0.903	1.786
0.807	3.319	1.026	0.964	0.884
0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000

Table 13.5
$x_1^*(ijk)$
k

i j	1	2	3	4	5
1 1	4.066	7.835	7.815	12.046	29.240
1 2	17.253	38.835	56.204	40.865	63.843
1 3	21.123	46.332	47.915	40.955	62.589
1 4	24.364	41.268	38.056	31.098	45.213
1 5	11.193	28.680	14.974	12.036	8.115
1 6	4	8	12	1	6
1 7	3	0	2	1	1
1 8	2	0	1	0	1
1 9	0	1	0	0	1
2 1	3.933	12.167	7.184	12.953	42.763
2 2	5.747	19.166	17.795	15.134	32.157
2 3	1.877	6.618	4.050	4.046	8.410
2 4	0.636	1.731	0.945	0.903	1.786
2 5	0.807	3.319	1.026	0.964	0.884
2 6	0	0	0	0	0
2 7	0	0	0	0	0
2 8	0	0	0	0	0
2 9	0	0	0	0	0

Table 13.6

$x_B^*(ijk) = x(ij.)x(.jk)/x(.j.)$ 2x9x5

i j	k = 1	2	3	4	5
1 1	[illegible]	[illegible]	6.536	10.893	31.371
1 2	[illegible]	[illegible]	52.092	39.421	67.579
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	1.000	1.000
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1.000
[illegible]	[illegible]	[illegible]	[illegible]	0.000	1.000
2 1	[illegible]	11.266	8.464	14.107	40.629
2 2	[illegible]	[illegible]	21.908	16.579	28.421
2 3	[illegible]	[illegible]	[illegible]	4.611	7.275
[illegible]	[illegible]	[illegible]	[illegible]	1.032	1.516
2 5	[illegible]	[illegible]	1.336	1.110	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	0.000	0.000
[illegible]	[illegible]	[illegible]	0.000	[illegible]	[illegible]
2 [illegible]	[illegible]	[illegible]	[illegible]	0.000	0.000
2 [illegible]	[illegible]	[illegible]	[illegible]	0.000	[illegible]

Table 13.7

$x_E(ijk)$

k = 1	2	3	4	5
3.48[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
16.[illegible]	38.[illegible]	[illegible]	[illegible]	[illegible]
20.643	47.[illegible]	[illegible]	[illegible]	[illegible]
21.[illegible]	41.[illegible]	[illegible]	[illegible]	[illegible]
10.970	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
3	0	2	1	1
2	0	[illegible]	[illegible]	[illegible]
0	1	0	0	1
4.514	11.266	[illegible]	[illegible]	[illegible]
6.800	16.[illegible]	[illegible]	[illegible]	[illegible]
2.35[illegible]	5.43[illegible]	[illegible]	[illegible]	7.275
0.806	1.3[illegible]	[illegible]	[illegible]	[illegible]
1.024	[illegible].732	1.33[illegible]	[illegible]	[illegible]
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

14. Acknowledgement

The research of the underlying theory was partially supported by the Air Force Office of Scientific Research, Office of Aerospace Research, U.S. Air Force, under Grant AFOSR-72-2348. The development and application of the techniques and methodology were partially supported by the Office of Naval Research under Contract Number N00014-67-A-0214 Task 0016, Project NR347-024 and by the Army Research Office, Office of Naval Research, and Air Force Office of Scientific Research by Contract No. N00014-67-A-0214-0015 (NR-042-267).

Computations were carried out at the Computer Center, The George Washington University. The support and encouragement of Professor Henry Solomon and Professor Herbert Solomon are gratefully acknowledged.

This paper was the basis of lectures presented by Professor S. Kullback 15 June 1973 at Osaka University, Osaka, Japan; 20 June 1973 at The Institute of Statistical Mathematics, Tokyo, Japan; and 6 July 1973 at Hiroshima University, Hiroshima, Japan. The opportunity and invitations to present this material were made possible by the partial support and interest of the Japan Society for the Promotion of Science, Professor Minoru Sakaguchi, Professor Kiyonori Kunisawa, Dr. Yukiyosi Kawada, Dr. Kameo Matusita, Professor F. Tanaka and Professor Sumiyasu Yamamoto. Their support and consideration are gratefully acknowledged.

The bibliography is essentially a compilation of those prepared by Dr. Marvin A. Kastenbaum and Dr. H.H. Ku covering the period through 1971 and permission to use their results is gratefully acknowledged.

15. Bibliography

The bibliography lists publications on contingency table analysis through 1972. Additional references to related topics may be found in the bibliographies contained in the books by D. R. Cox (1970) and H. O. Lancaster (1969). We will appreciate your calling our attention to possible references omitted from the bibliography.

REFERENCES

1972

ADAM, J. and ENKE, H. (1972). Analyse mehrdimensionaler kontingenztafeln mit hilfe des informationsmasses von Kullback. Biometrische Zeitschrift 14, 5, pp. 305-323.

BERKSON, J. (1972). Minimum discrimination information, the "no interaction" problem, and the logistic function. Biometrics 28, 2, pp. 443-468.

BRUNDEN, M. N. (1972). The analysis of non-independent 2x2 tables from 2xc tables using rank sums. Biometrics 28, 2, pp. 603-606.

CAUSEY, B. D. (1972). Sensitivity of raked contingency table totals to changes in problem conditions. Ann. Math. Statist. 43, 2, pp. 656-658.

COX, D. R. (1972). The analysis of multivariate binary data. Appl. Statist. 21, 2, pp. 113-120.

DARROCH, J. N. and RATCLIFF, D. (1972). Generalized iterative scaling for log-linear models. Ann. Math. Statist. 43, 5, pp. 1470-1480.

FIENBERG, S. E. (1972). The analysis of incomplete multiway contingency tables. Biometrics 28, 1, pp. 177-202.

GAIL, M. H. (1972). Mixed quasi-independence models for categorical data. Biometrics 28, 3, pp. 703-712.

GART, J. J. (1972). Interaction tests for 2xsxt contingency tables. Biometrika 59, 2, pp. 309-316.

GOKHALE, D. V. (1972). Analysis of log-linear models. J. Roy. Statist. Soc. Ser. B 34, 3, pp. 371-376.

GOODMAN, L. A. and KRUSKAL, W. H. (1972). Measures of association for cross-classifications, IV: simplification of asymptotic variances. J. Amer. Statist. Assoc. 67, pp. 415-421.

GRIZZLE, J. E. and WILLIAMS, O. D. (1972). Log-linear models and tests of independence for contingency tables. Biometrics 28, 1, pp. 137-156.

GRIZZLE, J. E. and WILLIAMS, O. D. (1972). Contingency tables having ordered response categories. J. Amer. Statist. Assoc. 67, pp. 55-63.

KOCH, G. G., IMREY, P. B., and REINFURT, D. W. (1972). Linear model analysis of categorical data with incomplete response vectors. Biometrics 28, 3, pp. 663-692.

NATHAN, G. (1972). Asymptotic power of tests for independence in contingency tables from stratified samples. J. Amer. Statist. Assoc. 67, pp. 917-920.

VICTOR, N. (1972). Zur klassifizierung mehrdimensionaler kontingenztafeln. Biometrics 28, 2, pp. 427-442.

1971

ALTHAM, P. M. E. (1971). Exact Bayesian analysis of the intraclass 2x2 table. Biometrika 58, 3, pp. 679-680.

ALTHAM, P. M. E. (1971). The analysis of matched proportions. Biometrika 58, 3, pp. 561-576.

BELLE, G. V. and CORNELL, R. G. (1971). Strengthening tests of symmetry in contingency tables. Biometrics 27, pp. 1074-1078.

BISHOP, Y. M. M. (1971). Effects of collapsing multidimensional contingency tables. Biometrics 27, pp. 545-562.

COHEN, J. E. (1971). Estimation and interaction in a censored 2x2x2 contingency table. Biometrics 27, 379-386.

DEMPSTER, A. P. (1971). An overview of multivariate data analysis. Journal Multivariate Analysis 1, 316-347.

FRYER, J. G. (1971). On the homogeneity of the marginal distributions of a multidimensional contingency table. J. Roy. Statist. Soc. Ser. A **134**, pp. 368-371.

GART, J. J. (1971). On the ordering of contingency tables for significance tests. Technometrics **13**, pp. 910-911.

GART, J. J. (1971). The comparison of proportions: a review of significance tests, confidence intervals, and adjustments for stratification. Rev. Inst. Internat. Statist. **29**, pp. 148-169.

GOKHALE, D. V. (1971). An iterative procedure for analysing log-linear models. Biometrics **27**, pp. 681-687.

GOODMAN, L. A. (1971). Partitioning of chi-square, analysis of marginal contingency tables, and estimation of expected frequencies in multidimensional contingency tables. J. Amer. Statist. Assoc. **66**, pp. 339-344.

GOODMAN, L. A. (1971). Some multiplicative models for the analysis of cross-classified data. Proc. 6th Berkeley Symp., Berkeley and Los Angeles, University of California Press.

GOODMAN, L. A. (1971). The analysis of multidimensional contingency tables: stepwise procedures and direct estimation methods for building models for multiple classifications. Technometrics **13**, pp. 33-61.

GRIZZLE, J. E. (1971). Multivariate logit analysis. Biometrics **27**, pp. 1057-1062.

JOHNSON, W. D., and KOCH, G. G. (1971). A note on the weighted least squares analysis of the Ries-Smith contingency table data. Technometrics **13**, pp. 438-447.

KOCH, G. G., IMREY, P. B., and REINFURT, D. W. (1971). Linear model analysis of categorical data with incomplete response vectors. Institute of Statistics Mimeo Series No. 790, University of North Carolina.

KOCH, G. G., JOHNSON, W. D., and TOLLEY, H. D. (1971). An application of linear models to analyze categorical data pertaining to the relationship between survival and extent of disease. Institute of Statistics Mimeo Series No. 776, University of North Carolina.

KOCH, G. G. and REINFURT, D. W. (1971). The analysis of categorical data from mixed models. Biometrics 27, pp. 157-173.

KU, H. H. (1971). Analysis of information - an alternative approach to the detection of a correlation between the sexes of adjacent sibs in human families. Biometrics 27, pp. 175-182.

KU, H. H., VARNER, R., and KULLBACK, S. (1971). On the analysis of multidimensional contingency tables. J. Amer. Statist. Assoc. 66, pp. 55-64.

KULLBACK, S. (1971). Marginal homogeneity of multidimensional contingency tables. Ann. Math. Statist. 42, pp. 594-606.

KULLBACK, S. (1971). The homogeneity of the sex ratio of adjacent sibs in human families. Biometrics 27, pp. 452-457.

NAM, J. (1971). On two tests for comparing matched proportions. Biometrics 27, pp. 945-959.

PEACOCK, P. B. (1971). The non-comparability of relative risks from different studies. Biometrics 27, pp. [illegible].

PERITZ, E. (1971). Estimating the ratio of two marginal probabilities in a contingency table. Biometrics 27, pp. 223-225.

RATCLIFF, D. (1971). Topics on independence and correlation for bounded sum variables. Ph.D. thesis, School of Mathematical Sciences, the Flinders University of South Australia, June 1971.

SIMON, G. A. (1971). Information distances and exponential families, with applications to contingency tables. Technical Report No. 32, November 26, 1971, Department of Statistics, Stanford University.

THOMAS, D. G. (1971). Exact confidence limits for the odds ratio in a 2x2 table. Appl. Statist. 20, pp. 105-110.

YASSAEE-MALEKI (YASSAEE), H. BAYAT (1971). On comparison of various estimators and their associated statistics in rxc and rxcx2 contingency tables. Ph.D. Dissertation, The George Washington University.

ZELEN, M. (1971). The analysis of several 2x2 contingency tables. Biometrika 58, pp. 129-137.

1970

ALTHAM, PATRICIA M. E. (1970). The measurement of association of rows and columns for an rxs contingency table. J. Roy. Statist. Soc. Ser. B 32, pp. 63-73.

BHAPKAR, V. P. (1970). Categorical data analysis of some multivariate tests. Essays in Probability and Statistics (R. C. Bose et al., eds.). The University of North Carolina Press, pp. 85-110.

CAMPBELL, L. L. (1970). Equivalence of Gauss's principle and minimum discrimination information estimation of probabilities. Ann. Math. Statist. 41, pp. 1011-1015.

COX, D. R. (1970). The Analysis of Binary Data. Methuen & Co., Ltd., London.

CRADDOCK, J. M. and FLOOD, C. R. (1970). The distribution of the chi-square statistic in small contingency tables. Appl. Statist. 19, pp. 173-181.

FIENBERG, S. E. (1970). Quasi-independence and maximum likelihood estimation in incomplete contingency tables. J. Amer. Statist. Assoc. 65, 332, pp. 1610-1616.

FIENBERG, S. E. (1970). An iterative procedure for estimation in contingency tables. Ann. Math. Statist. 41, pp. 907-917.

FIENBERG, S. E. (1970). The analysis of multidimensional contingency tables. Ecology 51, 2, pp. 419-433.

FIENBERG, S. E. and GILBERT, J. P. (1970). Geometry of a two by two contingency table. J. Amer. Statist. Assoc. 65, pp. 694-701.

FIENBERG, S. E. and HOLLAND, P. W. (1970). Methods for eliminating zero counts in contingency tables. Random Counts in Scientific Work (G. P. Patil, ed.). The Pennsylvania State University Press.

GOOD, I. J., GOVER, T. N., and MITCHELL, G. J. (1970). Exact distributions for χ-squared and for the likelihood-ratio statistic for the equiprobable multinomial distribution. J. Amer. Statist. Assoc. 65, pp. 267-283.

GOODMAN, L. A. (1970). The multivariate analysis of qualitative data: interaction among multiple classifications. J. Amer. Statist. Assoc. 65, pp. 226-256.

KASTENBAUM, M. A. (1970). A review of contingency tables. Essays in Probability and Statistics (R. C. Bose et al., eds.). The University of North Carolina Press, pp. 407-438.

KULLBACK, S. (1970). Various applications of minimum discrimination information estimation, particularly to problems of contingency table analysis. Proceedings of the Meeting on Information Measures, University of Waterloo, Ontario, Canada, April 10-14, 1970, I-33-I-66.

KULLBACK, S. (1970). Minimum discrimination information estimation and application. Proceedings of the Sixteenth Conference on the Design of Experiments in Army Research, Development and Testing, 21 October 1970. ARO-D Report 71-3, 1-38 Proceedings of the Conference.

MANTEL, N. (1970). Incomplete contingency tables. Biometrics 26, pp. 291-304.

MOLK, YEHUDA (1970). On estimation of probabilities in contingency tables with restrictions on marginals. Ph.D. dissertation, The George Washington University, February 1970.

ODOROFF, C. L. (1970). Minimum logit chi-square estimation and maximum likelihood estimation in contingency tables. J. Amer. Statist. Assoc. 65, 332, pp. 1617-1631.

WAGNER, S. S. (1970). The maximum-likelihood estimate for contingency tables with zero diagonal. J. Amer. Statist. Assoc. 65, 331, pp. 1362-1383.

1969

ALTHAM, PATRICIA M. E. (1969). Exact Bayesian analysis of a 2x2 contingency table and Fisher's "exact" significance test. J. Roy. Statist. Soc. Ser. B. 31, pp. 261-269.

ARGENTIERO, P. D. (1969). χ-squared statistic for goodness of fit test, its derivation and tables. NASA Technical Report, TR-R-313.

BISHOP, Y. M. M. (1969). Full contingency tables, logits, and split contingency tables. Biometrics 25, pp. 383-400.

BISHOP, Y. M. M. and FIENBERG, S. E. (1969). Incomplete two-dimensional contingency tables. Biometrics 25, pp. 119-128.

DEMPSTER, A. P. (1969). Some theory related to fitting exponential models. Research Report S-4, Department of Statistics, Harvard University.

FIENBERG, S. E. (1969). Preliminary graphical analysis and quasi-independence for two-way contingency tables. Appl. Statist. 18, pp. 153-168.

GOODMAN, L. A. (1969). On partition χ-squared and detecting partial association in the three-way contingency tables. J. Roy. Statist. Soc. Ser. B. 31, pp. 486-498.

GRIZZLE, J. E., STARMER, C. F., and KOCH, G. G. (1969). Analysis of categorical data by linear models. Biometrics 25, pp. 489-504.

HEALY, M. J. R. (1969). Exact tests of significance in contingency tables. Technometrics 11, pp. 393-395.

IRELAND, C. T., KU, H. H., and KULLBACK, S. (1969). Symmetry and marginal homogeneity of an rxr contingency table. J. Amer. Statist. Assoc. 64, pp. 1223-1341.

KOCH, G. G. (1969). The effect of non-sampling errors on measures of association in 2x2 contingency tables. J. Amer. Statist. Assoc. 64, pp. 852-863.

KU, H. H. and KULLBACK, S. (1969). Analysis of multidimensional contingency tables: an information theoretical approach. Contributed papers, 37th Session of the International Statistical Institute, pp. 156-158.

LANCASTER, H. O. (1969). Contingency tables of higher dimensions. Bulletin of the International Statistical Institute. 43, 1, pp. 143-151.

LANCASTER, H. O. (1969). The Chi-Squared Distribution. Wiley, New York.

NAGNUR, B. N. (1969). LAMST and the hypotheses of no three factor interaction in contingency tables. J. Amer. Statist. Assoc. 64, pp. 207-215.

PLACKETT, R. L. (1969). Multidimensional contingency tables. A Survey of Models and Methods, Bulletin of the International Statistical Institute. 43, 1, pp. 133-142.

1968

BENNETT, B. M. (1968). Notes on χ-squared tests for matched samples. J. Roy.Statist. Soc. Ser. B 30, pp. 368-370.

BERKSON, J. (1968). Application of minimum logit chi-squared estimate to a problem of Grizzle with a notation on the problem of no interaction. Biometrics 24, pp. 75-96.

BHAPKAR, V. P. (1968). On the analysis of contingency tables with a quantitative response. Biometrics 24, pp. 329-338.

BHAPKAR, V. P. and KOCH, G. G. (1968). Hypotheses of "no interaction" in multidimensional contingency tables. Technometrics 10, pp. 107-123.

BHAPKAR, V. P. and KOCH, G. G. (1968). On the hypotheses of "no interaction" in contingency tables. Biometrics 24, pp. 567-594.

FIENBERG, S. E. (1968). The geometry of an r×c contingency table. Ann. Math. Statist. 39, pp. 1186-1190.

GOODMAN, L. A. (1968). The analysis of cross-classified data: independence, quasi-independence, and interactions in contingency tables with or without missing entries. J. Amer. Statist. Assoc. 63, pp. 1091-1131.

HAMDAN, M. A. (1968). Optimum choice of classes for contingency tables. J. Amer. Statist. Assoc. 63, pp. 291-297.

IRELAND, C. T. and KULLBACK, S. (1968). Contingency tables with given marginals. Biometrika 55, pp. 179-188.

IRELAND, C. T. and KULLBACK, S. (1968). Minimum discrimination information estimation. Biometrics 24, pp. 707-71[illegible].

KU, H. H. and KULLBACK, S. (1968). Interaction in multid[illegible] contingency tables: an information theoretic app[illegible] J. [illegible] Nat. Bur. Standards Sect. B 72, pp. 159-199.

KU, H. H., VARNER, R., and KULLBACK, S. (1968). Analysis [illegible] sional contingency tables. Proceedings of the F[illegible] on the Design of Experiments in Army Research, Development and Testing. ARO-D Report 69-2.

KULLBACK, S. (1968). Information Theory and Statistics. Dover Pub., Inc., New York.

KULLBACK, S. (1968). Probability densities with given marginals. Ann. Math. Statist. 39, pp. 1236-1243.

LEYTON, M. K. (1968). Rapid calculation of exact probabilities for 2x3 contingency tables. Biometrics 24, pp. 714-717.

MATHIEU, JEAN-RENE and LAMBERT, E. (1968). Un test de l'identite des marges dun tableau de correlation. C. R. Acad. Sci. Paris 267, pp. 832-834.

MOSTELLER, F. (1968). Association and estimation in contingency tables. J. Amer. Statist. Assoc. 63, pp. 1-28.

SLAKTER, M. J. (1968). Accuracy of an approximation to the power of the chi-square goodness of fit test with small but equal expected frequencies. J. Amer. Statist. Assoc. 63, pp. 912-924.

SUGIURA, N. and OTAKE, M. (1968). Numerical comparison of improvised methods of testing in contingency tables with small frequencies. Ann. Inst. Statist. Math. 20, pp. 507-517.

1967

BENNETT, B. M. (1967). Tests of hypothesis concerning matched samples. J. Roy. Statist. Soc. Ser. B 29, pp. 468-474.

BISHOP, Y. M. M. (1967). Multidimensional contingency tables: cell estimates. Ph.D. dissertation, Harvard University.

BLOCH, D. A. and WATSON, G. S. (1967). A Bayesian study of the multinomial distribution. Ann. Math. Statist. 38, pp. 1423-1435.

COX, D. R. and LAUH, E. (1967). A note on the graphical analysis of multidimensional contingency tables. Technometrics 9, pp. 481-488.

GOOD, I. J. (1967). A Bayesian significance test for multinomial distributions. J. Roy. Statist. Soc. Ser. B 29, pp. 339-431.

1966

ARMITAGE, P. (1966). The chi-square test for heterogeneity of proportions after adjustment for stratification. J. Roy. Statist. Soc. Ser. B **28**, pp. 150-163.

BHAPKAR, V. P. (1966). A note on the equivalence of two test criteria for hypotheses in categorical data. J. Amer. Statist. Assoc. **61**, pp. 228-235.

BHAPKAR, V. P. (1966). Notes on analysis of categorical data. Institute of Statistics Mimeo Series No. 477, University of North Carolina.

BHAT, B. R. and KULKARNI, S. R. (1966). LAMP test of linear and loglinear hypotheses in multinomial experiments. J. Amer. Statist. Assoc. **61**, pp. 236-245.

COX, D. R. (1966). A simple example of a comparison involving quantal data. Biometrika **53**, pp. 215-220.

CRADDOCK, J. M. (1966). Testing the significance of a 3x3 contingency table. The Statistician **16**, pp. 87-94.

GABRIEL, K. R. (1966). Simultaneous test procedures for multiple comparison on categorical data. J. Amer. Statist. Assoc. **61**, pp. 1081-1096.

GART, J. J. (1966). Alternative analyses of contingency tables. J. Roy. Statist. Soc. Ser. B **28**, pp. 164-179.

GOOD, I. J. (1966). How to estimate probabilities. J. Inst. Math. Appl. **2**, pp. 364-383.

KULLBACK, S. and KHAIRAT, M. A. (1966). A note on minimum discrimination information. Ann. Math. Statist. **37**, pp. 279-280.

MANTEL, N. (1966). Models for complex contingency tables and polychotomous dosage response curves. Biometrics **22**, pp. 83-95.

ASANO, C. (1965). On estimating multinomial probabilities by pooling incomplete samples. Ann. Inst. Statist. Math. 17, pp. 1-14.

BHAPKAR, V. P. and KOCH, G. G. (1965). On the hypothesis of "no interaction" in three-dimensional contingency tables. Institute of Statistics Mimeo Series No. 440, University of North Carolina.

BHAPKAR, V. P. and KOCH, G. G. (1965). Hypothesis of no interaction in four-dimensional contingency tables. Institute of Statistics Mimeo Series No. 449, University of North Carolina.

BHAT, B. R. and NAGNUR, B. N. (1965). Locally asymptotically most stringent tests and Lagrangian multiplier tests of linear hypotheses. Biometrika 52, 3 and 4, pp. 459-468.

BIRCH, M. W. (1965). The detection of partial association II: the general case. J. Roy. Statist. Soc. Ser. B 27, pp. 111-124.

CAUSSINUS, H. (1965). Contribution a l'analyse statistique des tableaux de correlation. Ann. Fac. Sci. Univ. Toulouse 29, pp. 77-182.

GOOD, I. J. (1965). The Estimation of Probabilities: An Essay on Modern Bayesian Methods. Research Monograph, 30. The MIT Press, Cambridge, Massachusetts.

KASTENBAUM, M. A. (1965). Contingency tables: a review. MRC Technical Summary Report No. 596. Mathematical Research Center, The University of Wisconsin.

KATTI, S. K. and SASTRY, A. N. (1965). Biological examples of small expected frequencies and the chi-square test. Biometrics 21, pp. 49-54.

LANCASTER, H. O. and BROWN, T. A. I. (1965). Size of χ-squared test in the symmetrical multinomials. Austral. J. Statist. 7, p. 40.

LEWONTIN, R. C. and FELSENSTEIN (1965). The robustness of homogeneity tests in 2xn tables. Biometrics 21, pp. 19-33.

MOTE, V. L. and ANDERSON, R. L. (1965). An investigation of the effect of misclassification on the properties of chi-squared tests in the analysis of categorical data. Biometrika 52, pp. 95-109.

RADHAKRISHNA, S. (1965). Combination of results from several 2x2 contingency tables. Biometrics 21, pp. 86-98.

1964

ALLISON, H. E. (1964). Computational forms for chi-square. Amer. Statist. 18, 1, pp. 17-18.

BENNETT, B. M. and NAKAMURA, E. (1964). Tables for testing significance in a 2x3 contingency table. Technometrics 6, 4, pp. 439-458.

BROSS, I. D. J. (1964). Taking a covariable into account. J. Amer. Statist. Assoc. 59, 307, pp. 725-736.

CHEW, V. (1964). Application of the negative binomial distribution with probability of misclassification. Virginia Journal of Science 15, 1, pp. 34-40.

GOODMAN, L. A. (1964). Simultaneous confidence limits for cross-product ratios in contingency tables. J. Roy. Statist. Soc. Ser. B 26, 1, pp. 86-102.

GOODMAN, L. A. (1964). Simple methods for analyzing three-factor interaction in contingency tables. J. Roy. Statist. Soc. 59, pp. 319-352.

GOODMAN, L. A. (1964). Interactions in multidimensional contingency tables. Ann. Math. Statist. 35, 2, pp. 632-646.

GOODMAN, L. A. (1964). Simultaneous confidence intervals for contrasts among multinomial populations. Ann. Math. Statist. 35, 2, pp. 716-725.

HARKNESS, W. L. and KATZ, L. (1964). Comparison of the power functions for the test of independence in 2x2 contingency tables. Ann. Math. Statist. **35**, 3, pp. 1115-1127.

LINDLEY, D. V. (1964). The Bayesian analysis of contingency tables. Ann. Math. Statist. **35**, 4, pp. 1622-1643.

PLACKETT, R. L. (1964). The continuity correction in 2x2 tables. Biometrika **51**, Parts 3 and 4, pp. 327-338.

PUTTER, J. (1964). The χ^2 goodness-of-fit test for a class of cases of dependent observations. Biometrika **51**, pp. 250-252.

SOMERS, R. H. (1964). Simple measures of association for the triple dichotomy. J. Roy. Statist. Soc. Ser. A **127**, 3, pp. 409-415.

TALLIS, G. M. (1964). The use of models in the analysis of some classes of contingency tables. Biometrics **24**, 4, pp. 832-839.

1963

BENNETT, B. M. and NAKAMURA, E. (1963). Tables for testing significance in a 2x3 contingency table. Technometrics **5**, 4, pp. 501-511.

BIRCH, M. W. (1963). Maximum likelihood in three-way contingency tables. J. Roy. Statist. Soc. Ser. B **25**, 1, pp. 220-233.

DARROCH, J. N. and SILVEY, S. D. (1963). On testing more than one hypothesis. Ann. Math. Statist. **34**, 2, pp. 555-567.

DIAMOND, E. L. (1963). The limiting power of categorical data chi-square tests analogous to normal analysis of variance. Ann. Math. Statist. **34**, 4, pp. 1432-1441.

EDWARDS, A. W. F. (1963). The measure of association in a 2x2 table. J. Roy. Statist. Soc. Ser. A **126**, 1, pp. 109-114.

FELDMAN, S. E. and KLINGER, E. (1963). Short cut calculation of the Fisher-Yates exact test. Psychometrika **28**, 3, pp. 289-291.

GOLD, R. A. (1963). Tests auxiliary to χ^2 tests in a Markov chain. *Ann. Math. Statist.* 34, 1, pp. 56-74.

GOOD, I. J. (1963). Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables. *Ann. Math. Statist.* 34, 3, pp. 911-934.

GOODMAN, L. A. (1963). On methods for comparing contingency tables. *J. Roy. Statist. Soc. Ser. A* 126, 1, pp. 94-108.

GOODMAN, L. A. (1963). On Plackett's test for contingency table interactions. *J. Roy. Statist. Soc. Ser. B* 25, 1, pp. 179-188.

GOODMAN, L. A. and KRUSKAL, W. H. (1963). Measures of association for cross classification III: approximate sampling theory. *J. Amer. Statist. Assoc.* 58, pp. 310-364.

KU, H. H. (1963). A note on contingency tables involving zero frequencies and the 2I test. *Technometrics* 5, 3, pp. 398-400.

MANTEL, N. (1963). Chi-square tests with one degree of freedom: extensions of the Mantel-Haenszel procedure. *J. Amer. Statist. Assoc.* 58, pp. 690-700.

NEWELL, D. J. (1963). Misclassification in 2x2 tables. *Biometrics* 19, 1, pp. 187-188.

OKAMOTO, M. (1963). Chi-square statistic based on the pooled frequencies of several observations. *Biometrika* 50, pp. 524-528.

RIES, P. N. and SMITH, H. (1963). The use of chi-square for preference testing in multidimensional problems. *Chem. Eng. Prog. Symposium Series* 59, 42, pp. 39-43.

WALSH, J. E. (1963). Loss in test efficiency due to misclassification for 2x2 tables. *Biometrics* 19, 1, pp. 158-162.

1962

DALY, C. (1962). A simple test for trends in a contingency table. Biometrics 18, 1, pp. 114-119.

DARROCH, J. N. (1962). Interactions in multi-factor contingency tables. J. Roy. Statist. Soc. Ser. B 24, 1, pp. 251-263.

FISHER, SIR RONALD A. (1962). Confidence limits for a cross-product ratio. Austral. J. Statist. 4, 1, p. 41.

GART, J. J. (1962). Approximate confidence limits for relative risks. J. Roy. Statist. Soc. Ser. B 24, 2, pp. 454-463.

GART, J. J. (1962). On the combination of relative risks. Biometrics 18, 4, pp. 601-610.

KINCAID, W. M. (1962). The combination of 2xm contingency tables. Biometrics 18, 2, pp. 224-228.

KULLBACK, S. KUPPERMAN, M. and KU, H. H. (1962). An application of information theory to the analysis of contingency tables with a table of 2N ln N, N = 1(1)10,000. J. Res. Nat. Bur. Standards Sect. B 66, pp. 217-243.

KULLBACK, S., KUPPERMAN, M., and KU, H. H. (1962). Tests for contingency tables and Markov chains. Technometrics 4, 4, pp. 573-608.

LEWIS, B. N. (1962). On the analysis of interaction in multi-dimensional contingency tables. J. Roy. Statist. Soc. Ser. A 125, 1, pp. 88-117.

PLACKETT, R. L. (1962). A note on interactions in contingency tables. J. Roy. Statist. Soc. Ser. B 24, 1, pp. 162-166.

TALLIS, G. M. (1962). The maximum likelihood estimation of correlation from contingency tables. Biometrics 18, 3, pp. 342-353.

BERGER, A. (1961). On comparing intensities of association between two binary characteristics in two different populations. J. Amer. Statist. Assoc. 56, pp. 889-903.

BHAPKAR, V. P. (1961). Some tests for categorical data. Ann. Math. Statist. 32, 1, pp. 72-83.

BILLINGSLEY, P. (1961). Statistical Inference for Markov Processes. Statistical Research Monographs, 2, The University of Chicago Press.

CLARINGBOLD, P. J. (1961). The use of orthogonal polynomials in the partition of chi-square. Austral. J. Statist. 3, 2, pp. 48-63.

FRIEDLANDER, D. (1961). A technique for estimating a contingency table, given the marginal totals and some supplementary data. J. Roy. Statist. Soc. Ser. A 124, 3, pp. 412-420.

GREGORY, G. (1961). Contingency tables with a dependent classification. Austral. J. Statist. 3, 2, pp. 42-47.

GRIZZLE, J. E. (1961). A new method of testing hypotheses and estimating parameters for the logistic model. Biometrics 17, 3, pp. 372-385.

KENDALL, M. G. and STUART, A. (1961). The Advanced Theory of Statistics. 2, Charles Griffin and Company, London.

OKAMATO, M. and ISHII, G. (1961). Test of independence in intraclass 2x2 tables. Biometrika 48, pp. 181-190.

ROGOT, E. (1961). A note on measurement errors and detecting real differences. J. Amer. Statist. Assoc. 56, pp. 314-319.

SCHULL, W. J. (1961). Some problems of analysis of multi-factor tables. Bull. Inst. Internat. Statist. 28, Part 3, pp. 259-270.

YATES, F. (1961). Marginal percentages in multiway tables of quantal data with disproportionate frequencies. Biometrics 17, 1, pp. 1-9.

<u>1960</u>

BENNETT, B. M. and HSU, P. (1960). On the power function of the exact test for the 2x2 contingency table. <u>Biometrika</u> 47, pp. 393-398.

GRIDGEMAN, N. T. (1960). Card-matching experiments: a conspectus of theory. <u>J. Roy. Statist. Soc. Ser. A</u> 123, 1, pp. 45-49.

ISHII, G. (1960). Intraclass contingency tables. <u>Ann. Inst. Statist. Math.</u> 12, pp. 161-207; corrections, p. 279.

KASTENBAUM, M. A. (1960). A note on the additive partitioning of chi-square in contingency tables. <u>Biometrics</u> 16, 3, pp. 416-422.

KUPPERMAN, M. (1960). On comparing two observed frequency counts. <u>Appl. Statist.</u> 9, 1, pp. 37-42.

LANCASTER, H. O. (1960). On tests on independence in several dimensions. <u>J. Austral. Math. Soc.</u> 1, pp. 241-254.

ROBERTSON, W. H. (1960). Programming Fisher's exact method of comparing two percentages. <u>Technometrics</u> 2, 1, pp. 103-107.

<u>1959</u>

ANDERSON, R. L. (1959). Use of contingency tables in the analysis of consumer preference studies. <u>Biometrics</u> 15, 4, pp. 582-590.

CHAKRAVARTI, I. M. and RAO, C. R. (1959). Tables for some small sample tests of significance for Poisson distributions and 2x3 contingency tables. <u>Sankhya</u> 21, Parts 3 and 4, pp. 315-326.

GOODMAN, L. A. and KRUSKAL, W. H. (1959). Measures of association for cross classification II: further discussion and references. <u>J. Amer. Statist. Assoc.</u> 54, pp. 123-163.

HALDANE, J. B. S. (1959). The analysis of heterogeneity, I. <u>Sankhya</u> 21, Parts 3 and 4, pp. 209-216.

HOYT, C. J., KRISHNAIAH, P. R., and TORRANCE, E. P. (1959). Analysis of complex contingency data. *Journal of Experimental Education* **27**, pp. 187-194.

KASTENBAUM, M. A. and LAMPHIEAR, D. E. (1959). Calculation of chi-square to test the no three-factor interaction hypothesis. *Biometrics* **15**, 1, pp. 107-115.

KULLBACK, S. (1959). *Information Theory and Statistics*. John Wiley and Sons, New York.

KUPPERMAN, M. (1959). A rapid significance test for contingency tables. *Biometrics* **15**, 4, pp. 625-628.

NASS, C. A. G. (1959). The χ^2 test for small expectations in contingency tables, with special reference to accidents and absenteeism. *Biometrika* **46**, pp. 365-385.

SILVEY, S. D. (1959). The Lagrangian multiplier test. *Ann. Math. Statist.* **30**, 2, pp. 389-407.

SOMERS, R. H. (1959). The rank analogue of product-moment partial correlation and regression, with application to manifold, ordered contingency tables. *Biometrika* **46**, pp. 241-246

STEYN, H. S. (1959). On χ^2-tests for contingency tables of negative binomial type. *Statistica Neerlandica* **13**, pp. 433-444.

WEINER, I. B. (1959). A note of the use of Mood's likelihood ratio test for item analyses involving 2x2 tables with small samples. *Psychometrika* **24**, 4, pp. 371-372.

1958

BLALOCK, H. M., Jr. (1958). Probabilistic interpretations for the mean square contingency. *J. Amer. Statist. Assoc.* **53**, pp. 102-105.

KASTENBAUM, M. A. (1958). Estimation of relative frequencies of four sperm types in Drosophila melanogaster. *Biometrics* **14**, 2, pp. 223-228.

MITRA, S. K. (1958). On the limiting power function of the frequency chi-square test. *Ann. Math. Statist.* **29**, pp. 1221-1233.

SNEDECOR, G. W. (1958). Chi-square of Bartlett, Mood and Lancaster in a 2^3 contingency table. *Biometrics* **14**, 4, pp. 560-562 (Query).

1957

BROSS, I. D. J. and KASTEN, E. L. (1957). Rapid analysis of 2x2 tables. *J. Amer. Statist. Assoc.* **52**, pp. 18-28.

CORSTEN, L. C. A. (1957). Partition of experimental vectors connected with multinomial distributions. *Biometrics* **13**, 4, pp. 451-484.

EDWARDS, J. H. (1957). A note on the practical interpretation of 2x2 tables. *Brit. J. Prev. Soc. Med.* **11**, pp. 73-78.

LANCASTER, H. O. (1957). Some properties of the bivariate normal distribution considered in the form of a contingency table. *Biometrika* **44**, pp. 289-292.

MOTE, V. L. (1957). An investigation of the effect of misclassification of the chi-square tests in the analysis of categorical data. Unpublished Ph.D. dissertation, North Carolina State College, Raleigh, North Carolina (also Institute of Statistics Mimeo Series No. 182).

ROY, S. N. (1957). *Some Aspects of Multivariate Analysis.* John Wiley and Sons, New York.

SAKODA, J. M. and COHEN, B. H. (1957). Exact probabilities for contingency tables using binomial coefficients. *Psychometrika* **22**, 1, pp. 83-86.

WOOLF, B. (1957). The log likelihood ratio test (the G-test). Methods and tables for tests of heterogeneity in contingency tables. *Annals of Human Genetics* **21**, pp. 397-409.

1956

FISHMAN, J. A. (1956). A note on Jenkins' "Improved Method for Tetrachoric r." Psychometrika 20, 3, pp. 305.

GOOD, I. J. (1956). On the estimation of small frequencies in contingency tables. J. Roy. Statist. Soc. Ser. B 18, 1, pp. 113-124.

GRIDGEMAN, N. T. (1956). A tasting experiment. Appl. Statist. 5, 2, pp. 106-112.

LEANDER, E. K. and FINNEY, D. J. (1956). An extension of the use of the χ^2 test. Appl. Statist. 5, 2, pp. 132-136.

MAINLAND, D., HERRERA, L. and SUTCLIFFE, M. I. (1956). Statistical tables for use with binomial samples - contingency tests, confidence limits, and sample size estimates. New York University College of Medicine, New York.

ROY, S. N. and KASTENBAUM, M. A. (1956). On the hypothesis of no "interaction" in a multiway contingency table. Ann. Math. Statist. 27, 3, pp. 749-757.

ROY, S. N. and MITRA, S. K. (1956). An introduction to some non-parametric generalizations of analysis of variance and multivariate analysis. Biometrika 43, Parts 3 and 4, pp. 361-376.

WATSON, G. S. (1956). Missing and "mixed-up" frequencies in contingency tables. Biometrics 12, 1, pp. 47-50.

1955

ARMITAGE, P. (1955). Tests for linear trends in proportions and frequencies. Biometrics 11, 3, pp. 375-386.

ARMSEN, P. (1955). Tables for significance tests of 2x2 contingency tables. Biometrika 42, pp. 494-505.

COCHRAN, W. G. (1955). A test of a linear function of the deviations between observed and expected numbers. J. Amer. Statist. Assoc. 50, pp. 377-397.

HALDANE, J. B. S. (1955). Substitutes for χ^2. *Biometrika* **42**, pp. 265-266.

HALDANE, J. B. S. (1955). A problem in the significance of small numbers. *Biometrika* **42**, pp. 266-267.

HALDANE, J. B. S. (1955). The rapid calculation of χ^2 as a test of homogeneity from a 2xn table. *Biometrika* **42**, pp. 519-520.

JENKINS, W. L. (1955). An improved method for tetrachoric r. *Psychometrika* **20**, 3, pp. 253-258.

KASTENBAUM, M. A. (1955). Analysis of data in multiway contingency tables. Unpublished doctoral dissertation. North Carolina State College, October 1955.

LESLIE, P. H. (1955). A simple method of calculating the exact probability in 2x2 contingency tables with small marginal totals. *Biometrika* **42**, pp. 522-523.

MITRA, S. K. (1955). Contributions to the statistical analysis of categorical data. North Carolina Institute of Statistics Mimeograph Series No. 142, December 1955.

ROY, S. N. and KASTENBAUM, M. A. (1955). A generalization of analysis of variance and multivariate analysis to data based on frequencies in qualitative categorical or class intervals. North Carolina Institute of Statistics Mimeograph Series No. 131, June 1955.

ROY, S. N. and MITRA, S. K. (1955). An introduction to some non-parametric generalizations of analysis of variance and multivariate analysis. North Carolina Institute of Statistics Mimeograph Series No. 139, November 1955.

SEKAR, C. C., AGARWALA, S. P. and CHAKRABORTY, P. N. (1955). On the power function of a test of significance for the difference between two proportions. *Sankhya* **15**, Part 4, pp. 381-390.

STUART, A. (1955). A test of homogeneity of the marginal distributions in a two-way classification. Biometrika [illegible], pp. 412-416.

WOOLF, B. (1955). On estimating the relation between blood group and disease. Annals of Human Genetics 19, pp. 251-[illegible].

YATES, F. (1955). A note on the application of the combination of probabilities test to a set of 2x2 tables. Biometrika 42, pp. 401-411.

YATES, F. (1955). The use of transformations and maximum likelihood in the analysis of quantal experiments involving two treatments. Biometrika 42, pp. 382-403.

1954

BROSS, I. D. J. (1954). Misclassification in 2x2 tables. Biometrics 10, 4, pp. 478-486.

COCHRAN, W. G. (1954). Some methods for strengthening the common chi-square tests. Biometrics 10, 4, pp. 417-451.

DAWSON, R. B. (1954). A simplified expression for the variance of the χ^2 function on a contingency table. Biometrika 41, p. 280.

GOODMAN, L. A. and KRUSKAL, W. H. (1954). Measures of association for cross classification. J. Amer. Statist. Assoc. 49, pp. 732-764.

KIMBALL, A. W. (1954). Short-cut formulas for the exact partition of chi-square in contingency tables. Biometrics 10, 4, pp. 452-458.

MCGILL, W. J. (1954). Multivariate information transmission. Psychometrika 19, 2, pp. 97-116.

1952

COCHRAN, W. G. (1952). The χ^2 test of goodness of fit. Ann. Math. Statist. 23, 3, pp. 315-345.

DYKE, G. V. and PATTERSON, H. D. (1952). Analysis of factorial arrangements when the data are proportions. *Biometrics* **8**, pp. 1-12.

1951

LANCASTER, H. O. (1951). Complex contingency tables treated by the partition of chi square. *J. Roy. Statist. Soc. Ser. B* **13**, pp. 242-249.

SIMPSON, C. H. (1951). The interpretation of interaction in contingency tables. *J. Roy. Statist. Soc. Ser. B* **13**, pp. 238-241.

1950

TOCHER, K. D. (1950). Extension of the Neyman-Pearson theory of tests to discontinuous variates. *Biometrika* **37**, pp. 130-144.

1949

HSU, P. L. (1949). The limiting distributions of functions of sample means and application to testing hypotheses. *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability* (1945, 1946), University of California Press, Berkeley and Los Angeles.

IRWIN, J. O. (1949). A note on the subdivision of chi-square into components. *Biometrika* **36**, pp. 130-134.

LANCASTER, H. O. (1949). The derivation and partition of chi-square in certain discrete distributions. *Biometrika* **36**, pp. 117-129.

1948

YATES, F. (1948). The analysis of contingency tables with groupings based on quantitative characters. *Biometrika* **35**, pp. 176-181.

1947

BARNARD, G. A. (1947). Significance tests for 2x2 tables. *Biometrika* **34**, pp. 123-138.

PEARSON, E. S. (1947). The choice of statistical tests illustrated on the interpretation of data classed in a 2x2 table. *Biometrika* **34**, pp. 139-167.

1946

CRAMER, H. (1946). *Mathematical Methods of Statistics.* Princeton University Press, p. 424.

1945

NORTON, H. W. (1945). Calculation of chi-square for complex contingency tables. *J. Amer. Statist. Assoc.* **40**, pp. 251-258.

1937

HALDANE, J. B. S. (1937). The exact value of the moments of the distribution of χ^2 used as a test of goodness of fit, when expectations are small. *Biometrika* **29**, pp. 133-143.

1935

BARTLETT, M. S. (1935). Contingency table interactions. *J. Roy. Statist. Soc. Supplement* **2**, pp. 248-252.

WILKS, S. S. (1935). The likelihood test of independence in contingency tables. *Ann. Math. Statist.* **6**, pp. 190-196.

1934

FISHER, R. A. (1934). *Statistical Methods for Research Workers*, Edinburgh 5th and subsequent editions. Oliver and Boyd Ltd., Section 21.02.

YATES, F. (1934). Contingency tables involving small numbers and the χ^2 test. *J. Roy. Statist. Soc. Supplement 1* **1**, pp. 217-235.

1924

FISHER, R. A. (1924). The conditions under which chi-square measures the discrepancy between observation and hypothesis. J. Roy. Statist. Soc. **37**, pp. 442-450.

1922

FISHER, R. A. (1922). On the interpretation of chi-square from contingency tables, and the calculation of P. J. Roy. Statist. Soc. **85**, pp. 87-94.

1900

PEARSON, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philos. Mag., Series 5 **50**, pp. 157-172.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER 285	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) The Information in Contingency Tables - An Application of Information-Theoretic Concepts to the Analysis of Contingency Tables.		5. TYPE OF REPORT & PERIOD COVERED Technical Report.
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) C. T. Ireland and S. Kullback		8. CONTRACT OR GRANT NUMBER(s) N00014-76-C-0475
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, Calif. 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (NR-042-267)
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 436 Arlington, Virginia 22217		12. REPORT DATE August 4, 1976
		13. NUMBER OF PAGES 51
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release; Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

contingency tables
information theory
minimum discrimination information statistic

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

(see reverse side)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. The analysis of the information in contingency tables is an aspect of multivariate (multiple variates) analysis with particular application to qualitative or categorical as well as quantitative variables.

The analysis is concerned with counts in multiway cross-classifications or multiway contingency tables. Multiway contingency tables, or cross-classifications of vectors of discrete random variables, provide a useful approach to the analysis of multivariate discrete data.

The method of analysis presented will bring out the various interrelationships among the classificatory variables in a multiway cross-classification or contingency table in many dimensions.

The procedure is based on the Principle of Minimum Discrimination Information Estimation, associated statistics and Analysis of Information. General computer programs are available to provide the necessary results for inference. An analysis of a four-way contingency table is presented for illustration of these techniques.

#235